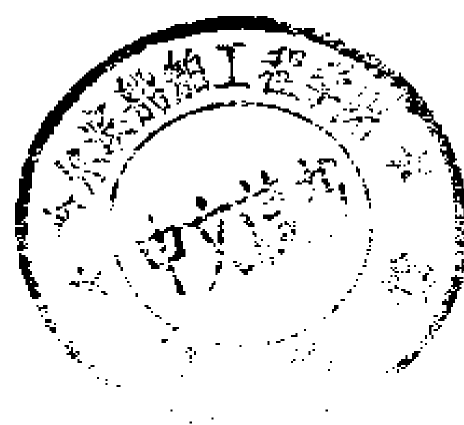
矩阵分析

杨克劭 包学游 编



哈尔滨工业大学出版社

（黑）新登字第4号

内容提要

本书是在已修《线性代数》课程的基础上，以工程和科技中常用的矩阵理论为选材标准而编写的。内容包括：线性空间与线性映射、酉空间与欧氏空间、矩阵的分解、向量与矩阵的范数、矩阵分析、矩阵函数和广义逆矩阵。书中各章均配有一定数量的习题。

本书为工科硕士研究生教材，也可供工程技术人员和科技工作者参考。

矩阵分析

杨克劭　包学游　编

*

哈尔滨工业大学出版社出版

新华书店首都发行所发行

黑龙江龙科印刷厂印刷

*

开本 787×1092　1/32　印张 8　字数 182 千字

1988 年 4 月第 1 版　1995 年 4 月第 2 次印刷

印数 3001—8000

ISBN 7-5603-0048-0/O·12　定价 7.80 元

前　　言

系统工程和自动控制已深入到现代工程技术的各个领域，使工程界对矩阵理论和知识的需求十分迫切。为了适应工科硕士研究生的教学需要，1982年由杨克劭编写了教材《矩阵分析》。在此基础上，几经修改和充实，重新编写了此书，使得教材更为恰当，结构更为合理。本书不仅注意到工程技术人员对矩阵理论的实际要求和学生的学时限制（三学分），也对数学本身的严谨性、系统性作了适当的考虑。

本书稿承蒙哈尔滨师范大学周汝奇副教授和张之凰副教授审阅，提出了很多宝贵意见和建议，并得到哈尔滨工业大学研究生院的大力支持，在此一并深表感谢。

限于编者学识水平，书中难免存在不当之处，热忱欢迎批评指正。

编　　者

1986年6月于哈尔滨工业大学

目　　录

第一章　线性空间与线性映射

线性空间是研究物理、力学中满足叠加原理的系统的数学模型，是对各种具体的线性运算封闭系统的共性加以概括、抽象而形成的新概念。线性映射则是用来研究线性空间之间关系的主要工具。因此，本章所讨论的内容既是已有线性代数知识的深化，也是本书的基础。本章主要介绍线性空间、线性子空间、线性映射和线性变换等基本概念和它们的性质。

§1.1　群、环、域的概念

为了今后能准确地叙述和理解有关的基本概念，学好矩阵分析，首先引入代数基本概念：群、环、域。

定义1　设G为非空集，若在G中定义了代数运算"$\circ$"，且满足下列条件时，称G关于代数运算"$\circ$"构成一个**群**(group)，简称群：

a)对任意的$\alpha,\beta\in G$，有唯一确定的元素$\alpha\circ\beta\in G$(自闭性、唯一性)；

b)$\alpha\circ(\beta\circ\gamma)=(\alpha\circ\beta)\circ\gamma\quad\forall\,\alpha,\beta,\gamma\in G$(结合律)；

c)G中存在元素e，使得

$$\alpha\circ e=e\circ\alpha=\alpha\quad\forall\,\alpha\in G$$

成立，并称e为G关于运算"$\circ$"的**单位元**；

d)对每个$\alpha\in G$，存在$\alpha^{-1}\in G$，使得

$$\alpha\circ\alpha^{-1}=\alpha^{-1}\circ\alpha=e$$

成立．并称 α^{-1} 为元素 α 关于运算"$\circ$"的**逆元素**。

如果群 G 中的元素关于运算"$\circ$"还满足交换律，即

$$\alpha\circ\beta=\beta\circ\alpha \quad \forall\ \alpha,\beta\in G$$

则称 G 为**交换群或阿贝尔(Abel)群**。

当群 G 是交换群、运算"$\circ$"称为"加法"时，则称群 G 为**加法群**，通常用"＋"代替"$\circ$"，并把加法单位元 e 称为**零元素**，记为"θ"，以别于数 0．称元素 α 的加法逆元素 α^{-1} 为 α 的**负元素**，记为"$-\alpha$"；若运算"$\circ$"称为"乘法"时，则称 G 为**乘法群**，通常用"$\cdot$"代替"$\circ$"，或在元素之间不写符号（例如 $\alpha\cdot\beta=\alpha\beta$）。

例 1 实数集 R 关于实数的加法运算构成加法交换群；在 R 中所有非零数之集，关于实数的乘法运算构成乘法交换群。

元素为实数的 $n\times n$ 阶矩阵的集合，记为 $R^{n\times n}$．关于矩阵的加法运算构成加法交换群，此时零元素为零矩阵 $O_{n\times n}$，矩阵 $A_{n\times n}$ 的负元素为 $-A$；在 $R^{n\times n}$ 中可逆矩阵组成的子集合，关于矩阵的乘法运算构成乘法群，此时单位元为 n 阶单位矩阵 I_n．可逆矩阵 A 的逆元素为其逆矩阵 A^{-1}．但此乘法群不是乘法交换群。

定义 2 在一个集合 $\tilde{R}$[①] 中定义两种代数运算"＋"、"$\cdot$"，且满足下列条件时，称 $\tilde{R}$ 关于这两种运算构成**环**(ring)：

a）$\tilde{R}$ 关于运算"＋"构成加法交换群；

b）对任意的 $\alpha,\beta\in\tilde{R}$ 有唯一确定的 $\alpha\cdot\beta\in\tilde{R}$（自闭性，唯

---

① 这里在 R 的上方加"~"，是由于习惯上已用 R 表示实数集，而又用 ring 的第一个字母的大写来代表环，故依此增区别。

一性)；

$c)\tilde{R}$ 关于乘法运算"·"结合律成立，即

$$\alpha\cdot(\beta\cdot\gamma)=(\alpha\cdot\beta)\cdot\gamma\quad\forall\,\alpha,\beta,\gamma\in\tilde{R}$$

$d)\tilde{R}$ 关于加法运算"+"与乘法运算"·"满足两个分配律，即

$$\alpha\cdot(\beta+\gamma)=\alpha\cdot\beta+\alpha\cdot\gamma\quad\forall\,\alpha,\beta,\gamma\in\tilde{R}$$

$$(\alpha+\beta)\cdot\gamma=\alpha\cdot\gamma+\beta\cdot\gamma\quad\forall\,\alpha,\beta,\gamma\in\tilde{R}$$

如果在环 $\tilde{R}$ 中的乘法运算"·"适合交换律，即

$$\alpha\cdot\beta=\beta\cdot\alpha\qquad\forall\,\alpha,\beta\in\tilde{R}$$

则称环 $\tilde{R}$ 为**交换环**。

例 2 全体整数构成的集合

$$Z=\{0,\pm1,\pm2,\cdots\}$$

关于实数的加法与乘法构成交换环；

数域 F 上的所有多项式的集合

$$P[x]=\{\sum_{i=0}^{m}a_ix^i\,|\,a_i\in F,m=0,1,2,\cdots\}$$

关于多项式的加法与乘法构成交换环；

设 $\tilde{R}$ 为交换环，则

$$\tilde{R}^{n\times n}=\{A=(a_{ij})\,|\,a_{ij}\in\tilde{R},i,j\in\underline{n}\}^{①}$$

关于矩阵的加法运算与乘法运算构成环。由于矩阵乘法是不能交换的，虽然 $\tilde{R}$ 是交换环，$\tilde{R}^{n\times n}$ 也只能是环而不能是交换环。

定义 3 一个至少包含两个元素的集合 F 满足下列条件时，称 F 为**域**(field)。

$a)F$ 是交换环；

① 记号 $i\in\underline{n}$ 表示 $i=1,2,\cdots,n$。

b)F 中所有不等于 θ 的元素对乘法构成交换群。

例 3 全体有理数集 Q、实数集 R、复数集 C 关于数的加法运算与乘法运算构成域。

例 4 设 p 为质数(即只能被 1 和它自身整除的大于 1 的自然数),在有限集

$$GF(p) = \{0,1,2,\cdots,p-1\}$$

中分别定义运算"$\circ$"、"$\cdot$"如下:

$$\alpha \circ \beta \triangleq \gamma_1 \qquad \forall\ \alpha,\beta \in GF(p)$$

$$\alpha \cdot \beta \triangleq \gamma_2 \qquad \forall\ \alpha,\beta \in GF(p)$$

其中 $\gamma_1 \equiv \alpha+\beta(\mathrm{mod}\,p)$,$\gamma_2 \equiv \alpha\beta(\mathrm{mod}\,p)$,$0 \leqslant \gamma_1,\gamma_2 < p$,即 $\alpha \circ \beta$ 为 $\alpha+\beta$ 被 p 除的余数,$\alpha \cdot \beta$ 为 $\alpha\beta$ 被 p 除的余数,不难证明 $GF(p)$ 是域。

特别,在 $GF(3)$ 上的 mod3 加法运算和乘法运算如下:

加 法

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

乘 法

	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

由此可见,$GF(3)$ 关于 mod3 加法运算和乘法运算构成域。

例 5 $\widetilde{R}^{n\times n}$ 是环但不是交换环,故不是域。

注意:每一个数学概念都有其适用范围,如讨论一个矩阵是否可逆,必须明确其元素取自什么集合,如设 $\widetilde{R}$ 是交换环

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \widetilde{R}^{2\times 2}$$

$$\Delta = a_{11}a_{22} - a_{12}a_{21} \in \widetilde{R}$$

$$\widetilde{A} = \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \in \widetilde{R}^{2\times 2}$$

容易验证

$$A\tilde{A} = \tilde{A}A = \Delta I_2$$

当且仅当 Δ 是 $\tilde{R}$ 的可逆元时，即 $\Delta^{-1}\in\tilde{R}$ 时，A 的逆矩阵 A^{-1} 才可由

$$A^{-1} = \Delta^{-1}\tilde{A} \in \tilde{R}^{2\times 2}$$

给出，亦即 A 是 $\tilde{R}^{2\times 2}$中可逆元的充要条件为 Δ 是 $\tilde{R}$ 中的可逆元。

一般，若 $A\in\tilde{R}^{n\times n}$为可逆元，则称 A 为么模阵。

设 F 为一个域，若 $A\in F^{n\times n}$，A 可逆的充要条件为 $\det A\neq 0$，此时称 A 为正则阵。

例如

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

作为 $R^{2\times 2}$的元素，由于 $\det A=2$，故 A 是正则阵，且

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$$

但将 A 作为 $Z^{2\times 2}$的元素，由于 2 在 Z 中无乘法逆元素，故 A 不是么模阵。又设

$$B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \in Z^{2\times 2}$$

$\det B=1$，且

$$B^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \in Z^{2\times 2}$$

故 B 既是 $Z^{2\times 2}$中的么模阵，也是 $R^{2\times 2}$中的正则阵。因此，"A 可逆的充要条件为 $\det A\neq 0$"，这只对域上的矩阵适用。由此可见，明确一个数学概念的适用范围是十分重要的。

§1.2 线性空间的定义与简单性质

在 n 维向量空间

$$R^n = \left\{ x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \middle| x_i \in R, i \in \underline{n} \right\}$$

或

$$C^n = \left\{ z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \middle| z_i \in C, i \in \underline{n} \right\}$$

中，关于向量的加法运算，数与向量的乘法运算所具有的性质．对于矩阵集合 $R^{m\times n}$ 或 $C^{m\times n}$、多项式集合 $P[x]$、齐次线性方程组 $A_{m\times n}x=\theta$ 的解集等的元素的加法与数乘运算，也具有相应的性质。因此，有必要对具有此类特性的集合的共性作进一步研究。为此我们引进线性空间的概念，并在本节讨论它的一些简单性质。

定义 1 当非空集合 V 与域 F 满足下列条件时，称 V 为域 F 上的**线性空间或向量空间**（并记为 $V(F)$）。

a）在 V 中定义了一个加法运算，记为"＋"，即对任意的 $x,y\in V$，有唯一确定的 $x+y\in V$，且 V 关于这个加法运算构成加法群；

b）在域 F 与集合 V 之间，定义一个数乘运算，对任意的 $\alpha\in F$，$x\in V$，有唯一确定的

$$\alpha x \in V$$

且适合

①$\alpha(\beta x)=(\alpha\beta)x \quad \forall \alpha,\beta\in F,\forall x\in V$

②对于域 F 的单位元 1，有

$$1x = x \qquad \forall x \in V$$

③下列分配律成立：

$$\alpha(x+y)=\alpha x+\alpha y \qquad \forall \alpha\in F,\forall x,y\in V$$

$$(\alpha+\beta)x=\alpha x+\beta x \qquad \forall \alpha,\beta\in F,\forall x\in V$$

例 1 按通常向量的加法与数乘运算，R^n 是实数域 R 上的线性空间；C^n 是复数域 C 上的线性空间；

$R^{n\times n}$ 关于矩阵的加法运算，域 R 中的数与 $R^{n\times n}$ 中矩阵的数乘运算，构成线性空间；

数域 F 上所有多项式的集合 $P[x]$，关于多项式的加法运算，数域 F 中的数与 $P[x]$ 中的多项式的数乘运算，构成线性空间；

区间 $[a,b]$ 上的连续函数集合（a、b 是两个固定的实数），关于函数的加法运算，域 R 上的数与函数的乘法运算，构成线性空间。

例 2 设正实数集

$$R^+=\{a\mid a>0,a\in R\}$$

对 R^+ 规定加法运算"$\dotplus$"：

$$a\dotplus b\triangleq ab \qquad \forall a,b\in R^+ \tag{1}$$

另外再规定 R 中的数与 R^+ 中的数的数乘运算"$\cdot$"：

$$\lambda\cdot\alpha\triangleq\alpha^\lambda \qquad \forall\lambda\in R,\forall\alpha\in R^+ \tag{2}$$

那么集合 R^+ 是域 R 上的线性空间。

证 由(1)，因为 $a>0,b>0$，故 $ab>0$，所以 $a\dotplus b=ab\in R^+$，且

$$a\dotplus b=ab=ba=b\dotplus a$$

$$(a\dotplus b)\dotplus c=ab\dotplus c=abc=a\dotplus(b\dotplus c)$$

$$\forall a,b,c \in R^+$$

在 R^+ 中存在唯一的"+"运算零元素 1，使得

$$a+1=1+a=1a=a \quad \forall a \in R^+$$

又由于 $a\in R^+$ 时，$a>0$，故 $\frac{1}{a}>0$，$\frac{1}{a}\in R^+$，且

$$a+\frac{1}{a}=\frac{1}{a}+a=\frac{1}{a}a=1 \quad \forall a \in R^+$$

即 R^+ 构成加群；

其次，由于 $a\in R^+$，显然

$$a^\lambda > 0 \qquad \forall \lambda \in R, \forall a \in R^+$$

故 $\lambda \cdot a=a^\lambda \in R^+$，且唯一，又

$$\lambda \cdot (\mu \cdot a) = (a^\mu)^\lambda = a^{\lambda\mu} = (\lambda\mu) \cdot a$$

$$\forall \lambda,\mu \in R, \forall a \in R^+$$

取 $1\in R$，有

$$1 \cdot a = a^1 = a \qquad \forall a \in R^+$$

且

$$\lambda \cdot (a+b) = \lambda \cdot (ab) = (ab)^\lambda = a^\lambda b^\lambda = \lambda \cdot a + \lambda \cdot b$$

$$\forall \lambda \in R, \forall a,b \in R^+$$

$$(\lambda+\mu) \cdot a = a^{\lambda+\mu} = a^\lambda a^\mu = a^\lambda + a^\mu = \lambda \cdot a + \mu \cdot a$$

$$\forall \lambda,\mu \in R, \forall a \in R^+$$

所以 R^+ 是域 R 上的线性空间(对于所规定的加法运算与数乘运算)。

注意：现行空间 $V(F)$ 与域 F 有密切关系，例如，R^n 是域 R 上的线性空间，但 R^n 不是域 C 上的线性空间。

由于线性空间 $V(F)$ 也可称为"向量空间"，故 $V(F)$ 的元素也可称为"向量"，当然，与 R^n 或 C^n 的元素相比，这里的"向量"是广义的向量，今后不再说明，而应作这样的理解。$x\in$

$V(F)$的**逆元素**$-x$也可称为x的**逆向量**，并可由此定义$V(F)$中向量之间的"减法运算"为

$$x-y \triangleq x+(-y) \qquad \forall\, x,y \in V(F)$$

线性空间具有以下简单性质：

性质 1 线性空间$V(F)$的零元素唯一；逆元素唯一。

事实上，设θ_1、θ_2均为$V(F)$的零元素，则

$$\theta_1 = \theta_1 + \theta_2 = \theta_2 + \theta_1 = \theta_2$$

其次，设x_1，x_2均为$x \in V(F)$的逆元素，θ为$V(F)$的零元素，则

$$\begin{aligned} x_1 &= x_1 + \theta = x_1 + (x + x_2) = (x_1 + x) + x_2 \\ &= \theta + x_2 = x_2 \end{aligned}$$

性质 2 设α，0，-1，$1 \in F$，x，$-x$，$\theta \in V(F)$，则

a) $0x=\theta$

b) $(-1)x=-x$

c) $\alpha\theta=\theta$

d) 若$\alpha x=\theta$，则$\alpha=0$或$x=\theta$。

事实上，因为

$$\begin{aligned} x + 0x &= 1x + 0x = (1+0)x \\ &= 1x = x \end{aligned}$$

故

$$0x=\theta$$

因为

$$\begin{aligned} x + (-1)x &= 1x + (-1)x = (1-1)x \\ &= 0x = \theta \end{aligned}$$

故

$$(-1)x = -x$$

因为

$$\alpha\theta = \alpha(x - x) = \alpha x - \alpha x$$
$$= (\alpha - \alpha)x = 0x = \theta$$

故

$$\alpha\theta = \theta$$

若 $\alpha \neq 0$ 且 $x \neq \theta$ 则

$$x = 1x = (\frac{1}{\alpha}\alpha)x = \frac{1}{\alpha}(\alpha x)$$
$$= \frac{1}{\alpha}\theta = \theta$$

这与 $x \neq \theta$ 矛盾，故 $\alpha \neq 0$ 与 $x \neq \theta$ 不能同时成立。

§1.3 线性空间的基

在 R^n 中有了向量的坐标表示式后，对于理论分析和实际应用都十分方便。为此，本节将 R^n 中有关基、维数和坐标等概念推广到一般线性空间中来。首先需要定义 $V(F)$ 中元素组的线性相关性等基本概念。

定义 1 设 $x_i \in V(F)$，$\alpha_i \in F$，$i \in \underline{m}$，则表达式

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m$$
$$= \sum_{i=1}^{m} \alpha_i x_i \qquad (x \in V(F)) \tag{1}$$

称为 x 用 x_1、x_2、…、x_m **线性表示**，或称 x 是 x_1、x_2、…、x_m 的**线性组合**。

定义 2 $V(F)$ 中的向量 x_1、x_2、…、x_m 所组成的向量组 $\{x_1、x_2、\cdots、x_m\}$ 称为是**线性相关的**，系指存在不全为零的 $\alpha_i \in F$，$i \in \underline{m}$，使

$$\sum_{i=1}^{m} \alpha_i x_i = \theta \tag{2}$$

否则称向量组 $\{x_1、x_2、\cdots、x_m\}$ 是**线性无关的**，即仅当

$$a_i = 0 \quad i \in \underline{m}$$

时，(2)成立。

与 R^n 类似，在 $V(F)$ 中，下列命题成立。

命题 1 $m \geqslant 2$ 时，$V(F)$ 中的向量组 $\{x_1, x_2, \cdots, x_m\}$ 线性相关的充要条件为，其中必有向量，它可由组中其余向量线性表示。

命题 2 若 $V(F)$ 中向量组的某一子向量组线性相关，则该向量组也线性相关。

与命题 2 等价的有下命题。

命题 3 若 $V(F)$ 中某向量组线性无关，则其任一子向量组也线性无关。

定义 3 线性空间 $V(F)$ 中的向量组 $\{x_1, x_2, \cdots, x_n\}$ 称为 $V(F)$ 的**基向量组**或**基**，系指

a) $\{x_1, x_2, \cdots, x_n\}$ 是线性无关组；

b) $V(F)$ 中任一向量 x，均可由向量组 $\{x_1、x_2、\cdots、x_n\}$ 线性表示。

线性空间 $V(F)$ 的基向量组 $\{x_1, x_2, \cdots, x_n\}$ 所含向量的个数 n，称为 $V(F)$ 的**维数**，记为

$$\dim V(F) = n$$

并称 $V(F)$ 为 **n 维线性空间**，可简记为 $V_n(F)$。特别，若对任何 $n \in Z^+$（Z^+ 表示正整数集），在 $V(F)$ 中均可找到 n 个线性无关的向量，则称 $V(F)$ 为**无限维线性空间**。只含零向量的线性空间的维数是 0。

定义 4 设 $\{x_1, x_2, \cdots, x_n\}$ 为 $V_n(F)$ 的基，则对 $x \in V_n(F)$，其唯一表示式

$$x = \sum_{i=1}^{n} \alpha_i x_i \qquad \alpha_i \in F, i \in \underline{n}$$

中的 α_1、α_2、…、α_n 称为向量 x 关于基〔$x_1,x_2,\cdots,x_n$〕的坐标。

例 1 在 R^n 中,向量组

$$e_1=\begin{bmatrix}1\\0\\0\\\vdots\\0\end{bmatrix},e_2=\begin{bmatrix}0\\1\\0\\\vdots\\0\end{bmatrix},\cdots,e_n=\begin{bmatrix}0\\0\\\vdots\\0\\1\end{bmatrix}$$

是线性无关的,它是 R^n 的基,容易验证向量组

$$e_1'=\begin{bmatrix}1\\1\\1\\\vdots\\1\end{bmatrix},e_2'=\begin{bmatrix}0\\1\\1\\\vdots\\1\end{bmatrix},\cdots,e_n'=\begin{bmatrix}0\\0\\\vdots\\0\\1\end{bmatrix}$$

也是线性无关的,故它也是 R^n 的基,由此可见,线性空间的基不唯一。

例 2 若用记号 $P[x]_n$ 表示次数低于 n 次的多项式集合,易证

$$1,x,x^2,\cdots,x^{n-1}$$

是 $P[x]_n$ 的基,若 $f(x)\in P[x]_n$,且

$$f(x)=a_0+a_1x+a_2x^2+\cdots+a_{n-1}x^{n-1}$$

则 a_0、a_1、a_2、…、a_{n-1}就是 $f(x)$关于基〔$1,x,x^2,\cdots,x^{n-1}$〕的坐标。又若取

$$1,(x-a),(x-a)^2,\cdots,(x-a)^{n-1}$$

为 $P[x]_n$ 的基,其中 a 为域 F 中常数,则由泰勒(Taylor)公式知

$$f(x)=f(a)+f'(a)(x-a)+\frac{f''(a)}{2!}(x-a)^2+\cdots$$

$$+\frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1}$$

故 $f(a), f'(a), \frac{f''(a)}{2!}, \cdots, \frac{f^{(n-1)}(a)}{(n-1)!}$ 为 $f(x)$ 关于基 $[1, (x-a), (x-a)^2, \cdots, (x-a)^{n-1}]$ 的坐标。

注意：$n(n\geqslant 1)$ 次多项式集合，关于多项式的加法运算、数与多项式的乘法运算是不能构成线性空间的。

在线性空间中，元素的坐标由基唯一确定，当基改变时，其坐标将随之改变，如果 $V_n(F)$ 的两个基之间的关系已知，如何由此推断 $V_n(F)$ 的元素关于这两个基的坐标之间的关系？

设 $[\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]$ 与 $[\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_n']$ 为 $Vn(F)$ 的两个基，且

$$\begin{cases} \varepsilon_1' = a_{11}\varepsilon_1 + a_{21}\varepsilon_2 + \cdots + a_{n1}\varepsilon_n \\ \varepsilon_2' = a_{12}\varepsilon_1 + a_{22}\varepsilon_2 + \cdots + a_{n2}\varepsilon_n \\ \qquad \cdots \quad \cdots \quad \cdots \quad \cdots \\ \varepsilon_n' = a_{1n}\varepsilon_1 + a_{2n}\varepsilon_2 + \cdots + a_{nn}\varepsilon_n \end{cases} \tag{3}$$

若记

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

则(3)式可表为

$$[\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_n'] = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]A \tag{4}$$

或

$$\begin{bmatrix} \varepsilon_1' \\ \varepsilon_2' \\ \vdots \\ \vdots \\ \varepsilon_n' \end{bmatrix} = A^T \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{5}$$

其中矩阵 A 称为由基 $[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n]$ 到基 $[\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n']$ 的**过渡矩阵**或**基变换矩阵**，它是正则的。

定理 设 $[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n]$ 与 $[\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n']$ 为 $V_n(F)$ 的两个基，且

$$[\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n'] = [\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n]A$$

若 $x\in V_n(F)$，有

$$\begin{aligned} x &= \alpha_1\varepsilon_1+\alpha_2\varepsilon_2+\cdots+\alpha_n\varepsilon_n \\ &= \alpha_1'\varepsilon_1'+\alpha_2'\varepsilon_2'+\cdots+\alpha_n'\varepsilon_n' \end{aligned} \tag{6}$$

则

$$\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \vdots\\ \alpha_n\end{bmatrix} = A\begin{bmatrix}\alpha_1'\\ \alpha_2'\\ \vdots\\ \vdots\\ \alpha_n'\end{bmatrix} \tag{7}$$

或

$$\begin{bmatrix}\alpha_1'\\ \alpha_2'\\ \vdots\\ \vdots\\ \alpha_n'\end{bmatrix} = A^{-1}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \vdots\\ \alpha_n\end{bmatrix} \tag{8}$$

证 由 x 的表示式(6)有

$$x = [\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n]\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \vdots\\ \alpha_n\end{bmatrix}$$

$$
= [\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_n'] \begin{bmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \vdots \\ \alpha_n' \end{bmatrix}
$$

$$
= [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n] A \begin{bmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \vdots \\ \alpha_n' \end{bmatrix}
$$

再由坐标表示的唯一性，即得(7)。又 A 可逆，故也可得(8)式。

(7)式或(8)式就是当基变换矩阵已知时，$\boldsymbol{x}$ 关于两个基的坐标之间的关系式。

例 3 设$[\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n]$与$[\boldsymbol{e}_1', \boldsymbol{e}_2', \cdots, \boldsymbol{e}_n']$为 R^n 的两个基(见例 1)，若 $x \in R^n$ 且

$$
\begin{aligned}
\boldsymbol{x} &= \alpha_1 \boldsymbol{e}_1 + \alpha_2 \boldsymbol{e}_2 + \cdots + \alpha_n \boldsymbol{e}_n \\
&= \alpha_1' \boldsymbol{e}_1' + \alpha_2' \boldsymbol{e}_2' + \cdots + \alpha_n' \boldsymbol{e}_n'
\end{aligned}
$$

求过渡矩阵 A 及 $\boldsymbol{x}$ 的两个坐标之间的关系式。

解 显然有

$$
[\boldsymbol{e}_1', \boldsymbol{e}_2', \cdots, \boldsymbol{e}_n'] = [\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n] \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 1 & \cdots & \cdots & 1 \end{bmatrix}
$$

即过渡矩阵

$$A=\begin{bmatrix}1 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 1 & \cdots & \cdots & 1\end{bmatrix}$$

容易求得

$$A^{-1}=\begin{bmatrix}1 & 0 & \cdots & \cdots & 0\\ -1 & \ddots & \ddots & \ddots & \vdots\\ 0 & \ddots & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & -1 & 1\end{bmatrix}$$

代入(8)式有

$$\alpha_1{}'=\alpha_1,\alpha_2{}'=\alpha_2-\alpha_1,\alpha_3{}'=\alpha_3-\alpha_2,\cdots,$$
$$\alpha_n{}'=\alpha_n-\alpha_{n-1}$$

例 4 $R^{m\times n}$是域 R 上的 mn 维线性空间。因为若令

$$E_{ij}=(e_{kl}{}^{(ij)})_{m\times n}\qquad i,k\in \underline{m},j,l\in \underline{n}$$

其中

$$e_{kl}{}^{(ij)}=\begin{cases}1 & k=i \text{ 且 } l=j\\ 0 & \text{其余}\end{cases}$$

易证 $E_{ij}(i\in \underline{m},j\in \underline{n})$是 $R^{m\times n}$中的一个基，设

$$A=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}\in R^{m\times n}$$

则

$$A=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}E_{ij}$$

即 $a_{ij}(i\in \underline{m},j\in \underline{n})$为 A 关于基 $E_{ij}(i\in \underline{m},j\in \underline{n})$的坐标。

§1.4 线性子空间

线性齐次方程组

$$Ax=\theta \qquad A\in R^{m\times n}, x\in R^n$$

的解的全体组成的集合是 R^n 的子集，它关于 R^n 的加法运算与数乘运算也构成线性空间；$P[x]_n$ 是 $P[x]$ 的子集，它关于多项式的加法运算、数与多项式的乘法运算也构成线性空间。线性空间的子集，其本身在原线性空间的加法运算与数乘运算下，仍构成线性空间的这一特性是十分重要的，它对于深入认识线性空间的性质、线性空间的分解与构造都是必不可少的。

定义 1 若线性空间 $V(F)$ 的非空子集 W 关于 $V(F)$ 中的加法运算与数乘运算也构成线性空间，则称 W 为 $V(F)$ 的**线性子空间**，简称**子空间**，记为 $W(F)$。

例 1 易证 R^n 的子集

$$S=\left\{x \,\middle|\, x=\begin{bmatrix}0\\x_2\\ \vdots\\ \vdots\\ x_n\end{bmatrix}\in R^n\right\}$$

关于 R^n 中向量的加法运算及数与向量的乘法运算构成线性空间，故 $S(R)$ 是 R^n 的线性子空间，且不难证明

$$\dim S(R)=n-1$$

例 2 $R^n\subset C^n$，且 R^n 与 C^n 均为线性空间，但 R^n 不是 C^n 的线性子空间，因为 R^n 与 C^n 作为线性空间的数域是不同的。

$V(F)$ 的零元素 θ 所组成的单元素集是 $V(F)$ 在域 F 上的零维线性子空间，并称它为 $V(F)$ 的**零子空间**；$V(F)$ 本身

也是 $V(F)$ 的线性子空间。$V(F)$ 的这两个特殊线性子空间，称为 $V(F)$ 的**平凡子空间**。除此以外，$V(F)$ 的其它线性子空间称为 $V(F)$ 的**非平凡子空间**。又 $V(F)$ 中除去自身以外的其它子空间，由于均由 $V(F)$ 的真子集构成，故称它们为 $V(F)$ 的**真子空间**。

欲证 $V(F)$ 的一个子集为其线性子空间，只须证明这个子集关于 $V(F)$ 的加法运算和数乘运算封闭即可。

例 3 设 $[x_1, x_2, \cdots, x_r]$ 为 $V(F)$ 的一个向量组，易证

$$T = \left\{ x \middle| x = \sum_{i=1}^{r} \alpha_i x_i, \alpha_i \in F, i \in \underline{r} \right\}$$

为 $V(F)$ 的子空间，并称它为 $V(F)$ 的由 $[x_1, x_2, \cdots, x_r]$ **生成的子空间**，简记为

$$T = \operatorname{span}[x_1, x_2, \cdots, x_r]$$

设 $[x_1, x_2, \cdots, x_n]$ 为 $V(F)$ 的一个向量组，其**极大无关部分组**（简称**极大无关组**）中所含向量的个数，称为向量组 $[x_1, x_2, \cdots, x_n]$ 的**秩**，记为 $\operatorname{rank}[x_1, x_2, \cdots, x_n]$。

线性子空间有以下性质：

定理 1 设 $x_i \in V(F), i \in \underline{r}$，则

$$\dim \operatorname{span}[x_1, x_2, \cdots, x_r] = \operatorname{rank}[x_1, x_2, \cdots, x_r] \qquad (1)$$

证 不失一般性，设 $x_1, x_2, \cdots, x_m (m \leqslant r)$ 为 $[x_1, x_2, \cdots, x_r]$ 的一个极大无关组，则

$$\operatorname{rank}[x_1, x_2, \cdots, x_r] = m$$

且对任意 $x \in \operatorname{span}[x_1, x_2, \cdots, x_r]$，有 $\alpha_i \in F, i \in \underline{r}$，使

$$x = \sum_{i=1}^{r} \alpha_i x_i = \sum_{i=1}^{m} \alpha_i x_i + \sum_{j=m+1}^{r} \alpha_j x_j$$

$$= \sum_{i=1}^{m} \alpha_i x_i + \sum_{j=m+1}^{r} \alpha_j \left(\sum_{k=1}^{m} \beta_{kj} x_k\right)$$

$$= \sum_{i=1}^{r}[\alpha_i + \sum_{j=m+1}^{r} \alpha_j\beta_{ij}]x_i$$

由此可知，向量组$[x_1,x_2,\cdots,x_m]$是$\text{span}[x_1,x_2,\cdots,x_r]$的一个基，故

$$\dim \text{span}[x_1,x_2,\cdots,x_r] = m$$

从而(1)式成立。证毕。

设$[x_1,x_2,\cdots,x_r]$与$[y_1,y_2,\cdots,y_s]$为$V(F)$的两个向量组，如果这两个向量组中的任一向量均可用另一向量组线性表示，则称这两个向量组是等价的。

定理 2 $\text{span}[x_1,x_2,\cdots,x_r]=\text{span}[y_1,y_2,\cdots,y_s]$的充要条件为向量组$[x_1,x_2,\cdots,x_r]$与$[y_1,y_2,\cdots,y_s]$等价。

证 必要性：设$\text{span}[x_1,x_2,\cdots,x_r]=\text{span}[y_1,y_2,\cdots,y_s]$，则任意$x_i\in\text{span}[x_1,x_2,\cdots,x_r]$，$i\in\underline{r}$，有$x_i\in\text{span}[y_1,y_2,\cdots,y_s]$，即$x_i$，$i\in\underline{r}$可由$y_1,y_2,\cdots,y_s$线性表示。同理可证$y_j$，$j\in\underline{s}$可由$x_1,x_2,\cdots,x_r$线性表示，故$[x_1,x_2,\cdots,x_r]$与$[y_1,y_2,\cdots,y_s]$等价；

充分性：设向量组$[x_1,x_2,\cdots,x_r]$与$[y_1,y_2,\cdots,y_s]$等价，则

$$x_i = \sum_{j=1}^{s}\alpha_{ij}y_j \qquad i\in\underline{r}$$

于是当$x\in\text{span}[x_1,x_2,\cdots,x_r]$时，有

$$x = \sum_{i=1}^{r}\beta_i x_i = \sum_{i=1}^{r}\beta_i(\sum_{j=1}^{s}\alpha_{ij}y_j)$$

$$= \sum_{j=1}^{s}(\sum_{i=1}^{r}\beta_i\alpha_{ij})y_j$$

故$x\in\text{span}[y_1,y_2,\cdots,y_s]$，从而有

$$\text{span}[x_1,x_2,\cdots,x_r] \subset \text{span}[y_1,y_2,\cdots,y_s]$$

同理可证

$$\text{span}[y_1,y_2,\cdots,y_t]\subset\text{span}[x_1,x_2,\cdots,x_r]$$

于是 $\text{span}[x_1,x_2,\cdots,x_r]=\text{span}[y_1,y_2,\cdots,y_t]$。

定理 3 设 $W_m(F)$是 $V_n(F)$的子空间，$[x_1,x_2,\cdots,x_m]$是 $W_m(F)$的一个基，则这个基定可扩充为 $V_n(F)$的基，即在 $V_n(F)$中一定能找到这样的 $n-m$ 个向量 x_{m+1}、x_{m+2}、$\cdots$、x_n，使 x_1、x_2、$\cdots$、x_m、x_{m+1}、$\cdots$、x_n 是 $V_n(F)$的基。

此定理简称为**基的扩充定理**。

证 对维数差 $n-m$，用数学归纳法证。

设 $n-m=0$，即 $n=m$，定理显然成立；

设 $n-m=k>0$ 时定理成立，证明 $n-m=k+1$ 时定理亦成立。若当 $n-m=k+1$ 时，x_1、x_2、$\cdots$、x_m 不是 $V_n(F)$的基，则在 $V_n(F)$中至少有一个向量不能被 x_1、x_2、$\cdots$、x_m 线性表示，不妨设此向量为 x_{m+1}于是向量组$[x_1,x_2,\cdots,x_m,x_{m+1}]$必定线性无关，由定理一知

$$\dim\text{span}[x_1,x_2,\cdots,x_m,x_{m+1}]=m+1$$

又因此时 $n-(m+1)=(n-m)-1=k+1-1=k$，故由归纳法假设知，子空间 $\text{span}[x_1,x_2,\cdots,x_m,x_{m+1}]$的基 x_1、x_2、$\cdots$、x_m、x_{m+1}可以扩充为 $V_n(F)$的基，因此，$W_m(F)$的基 x_1、x_2、$\cdots$、x_m 可以扩充为 $V_n(F)$的基。

下面定义子空间之间的"运算"，并由此讨论子空间之间的关系。

定义 2 设 $U(F)$与 $W(F)$均为线性空间 $V(F)$的子空间，称

$$U(F)\cap W(F)=\{x\mid x\in U(F)\text{ 且 }x\in W(F)\}$$

为 $U(F)$与 $W(F)$的交；称

$$U(F)+W(F)=\{x\mid x=u+w,u\in U(F),w\in W(F)\}$$

为 $U(F)$与 $W(F)$的和；特别，当 $U(F)\cap W(F)=\{\theta\}$时，和

$U(F)+W(F)$称为$U(F)$与$W(F)$的**直和**，为区别起见，直和记为$U(F)\oplus W(F)$。

注意：这里子空间的交与集合的交的概念是一致的，但子空间的和与集合的并的概念是不一致的。

例 4 在R^3中，设其子空间$U(R)$为xy坐标面，$W(R)$为zx坐标面，则

$$U(R)\cap W(R)=x\text{轴}$$

$$U(R)+W(R)=R^3$$

$$U(R)\cup W(R)=\{xy\text{坐标面与}zx\text{坐标面上的点}\}$$

定理 4 设$U(F)$与$W(F)$均为线性空间$V(F)$的子空间，则$U(F)\cap W(F)$与$U(F)+W(F)$亦为$V(F)$的子空间。

这个定理的证明留给读者自己练习。

子空间的运算"交"与"和"，易证其各自的交换律与结合律均成立，但"交"与"和"构成的"分配律"一般不成立。

例 5 设在R^2中，x_1、x_2是两个线性无关的向量，令

$$Q=\text{span}[x_1],S=\text{span}[x_2],T=\text{span}(x_1+x_2)$$

则

$$Q\cap(S+T)=Q\cap R^2=Q$$

$$(Q\cap S)+(Q\cap T)=\{\theta\}+\{\theta\}=\{\theta\}$$

由此可见

$$Q\cap(S+T)\neq(Q\cap S)+(Q\cap T)$$

定理 5 设$U(F)$与$W(F)$均为线性空间$V(F)$的子空间，则下列诸条件等价：

a) $U(F)+W(F)$是直和；

b) $z\in U(F)+W(F)$的表达式唯一；

c) 若x_1、x_2、…、x_r是$U(F)$的基，y_1、y_2、…、y_t是$W(F)$的基，则x_1、x_2、…、x_r、y_1、y_2、…、y_t是$U(F)+W(F)$的基；

$d)\dim(U(F)+W(F))=\dim U(F)+\dim W(F)$。

证 $a)\Rightarrow b)$，设 $U(F)+W(F)=U(F)\oplus W(F)$，则 $U(F)\cap W(F)=\{\theta\}$，若 $z\in U(F)+W(F)$ 的表达式不唯一，即设

$$z=x+y=x'+y',x,x'\in U(F),y,y'\in W(F)$$

且 $x\neq x'$（从而 $y\neq y'$）。于是

$$\theta\neq u=x-x'=y'-y\in U(F)\cap W(F)$$

这与 $U(F)\cap W(F)=\{\theta\}$ 矛盾，故 $b)$ 成立；

$b)\Rightarrow c)$，显然 $U(F)+W(F)=\text{span}[x_1,x_2,\cdots,x_r,y_1,y_2,\cdots,y_s]$，故只须证向量组 $x_1,x_2,\cdots,x_r,y_1,y_2,\cdots,y_s$ 线性无关，否则，则存在 $a\in R^r,b\in R^s$ 不全为零向量，使

$$[x_1,x_2,\cdots,x_r]a+[y_1,y_2,\cdots,y_s]b=\theta$$

于是 θ 有两种不同表达式，与 $b)$ 矛盾，故 $c)$ 成立；

$c)\Rightarrow d)$，显然；

$d)\Rightarrow a)$，反证。设 $U(F)\cap W(F)\neq\{\theta\}$，则取 $\theta\neq u\in U(F)\cap W(F)$，有不全为零的 $a_i,b_j\in F,i\in \underline{r},j\in \underline{s}$，使得

$$u=\sum_{i=1}^{r}a_ix_i=\sum_{j=1}^{s}b_jy_j$$

于是 $x_1,x_2,\cdots,x_r,y_1,y_2,\cdots,y_s$ 线性相关，所以

$$\dim(U(F)+W(F))<r+s=\dim U(F)+\dim W(F)$$

与 $d)$ 矛盾，故必有 $U(F)\cap W(F)=\{\theta\}$，即 $a)$ 成立。

定理 6 设 $U(F)$ 与 $W(F)$ 均为线性空间 $V(F)$ 的子空间，则

$$\dim U(F)+\dim W(F)=\dim[U(F)+W(F)]+\dim[U(F)\cap W(F)]$$

这个定理称为**子空间的维数定理**。

证 设 $\dim U(F)=r,\dim W(F)=s$，且 x_1、x_2、$\cdots$、x_m 为 $U(F)\cap W(F)$ 的基，易知 $U(F)\cap W(F)$ 分别为 $U(F)$ 与

$W(F)$的子空间，故由基的扩充定理知，可由 $\boldsymbol{x}_1$、$\boldsymbol{x}_2$、…、$\boldsymbol{x}_m$ 扩充为$U(F)$的基

$$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m,\xi_{m+1},\cdots,\xi_r$$

及 $W(F)$的基

$$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m,\eta_{m+1},\cdots,\eta_s$$

于是只须证

$$\dim[U(F)+W(F)]=r+s-m$$

显然

$$U(F)+W(F)=\mathrm{span}[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m,\xi_{m+1},\cdots,\xi_r,\eta_{m+1},\cdots,\eta_s]$$

故只须证向量组

$$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m,\xi_{m+1},\cdots,\xi_r,\eta_{m+1},\cdots,\eta_s \qquad (2)$$

线性无关，为此，设有等式

$$\alpha_1\boldsymbol{x}_1+\cdots+\alpha_m\boldsymbol{x}_m+\beta_{m+1}\xi_{m+1}+\cdots+\beta_r\xi_r+r_{m+1}\eta_{m+1}+\cdots+r_s\eta_s=\theta \qquad (3)$$

令

$$\boldsymbol{u}=\alpha_1\boldsymbol{x}_1+\cdots+\alpha_m\boldsymbol{x}_m+\beta_{m+1}\xi_{m+1}+\cdots+\beta_r\xi_r=-r_{m+1}\eta_{m+1}-\cdots-r_s\eta_s \qquad (4)$$

显然 $\boldsymbol{u}\in U(F)$且 $\boldsymbol{u}\in W(F)$，于是 $\boldsymbol{u}\in U(F)\cap W(F)$，故有

$$\boldsymbol{u}=\delta_1\boldsymbol{x}_1+\delta_2\boldsymbol{x}_2\cdots+\delta_m\boldsymbol{x}_m$$

代入(4)式有

$$\delta_1\boldsymbol{x}_1+\delta_2\boldsymbol{x}_2\cdots+\delta_m\boldsymbol{x}_m+r_{m+1}\eta_{m+1}+\cdots+r_s\eta_s=\theta$$

由于 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m,\eta_{m+1},\cdots,\eta_s$ 线性无关，故

$$\delta_1=\delta_2=\cdots=\delta_m=r_{m+1}=\cdots=r_s=0$$

将此结果代入(3)式得

$$\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2\cdots+\alpha_m\boldsymbol{x}_m-\beta_{m+1}\xi_{m+1}+\cdots+\beta_r\xi_r=\theta$$

又由于 $x_1,x_2,\cdots,x_m,\xi_{m+1},\cdots,\xi_r$ 线性无关，故

$$\alpha_1=\alpha_2=\cdots=\alpha_m=\beta_{m+1}=\cdots=\beta_r=0$$

因此，向量组(2)是线性无关的。

推论 设 $U(F)(\neq\{\theta\})$ 是线性空间 $V(F)$ 的子空间，则一定存在 $V(F)$ 的另一个空间 $W(F)$，使得

$$V(F)=U(F)\oplus W(F)$$

证明留给读者自己练习。

推论表明线性空间 $V(F)$ 可作直和分解，易知这种直和分解不是唯一的。

§1.5 线性映射与线性变换

为了研究线性空间之间的关系，引进线性映射的概念，而线性变换是线性映射的一特例，它是很有用的。

定义 1 设 M 与 M' 为两个集合，对于每个 $x\in M$，如果根据某种法则 σ，在 M' 中有唯一确定的 x' 与之对应，那么称 σ 为由 M 到 M' 的一个映射。记为

$$\sigma: M\to M'$$

或

$$\sigma(x)=x'$$

此时 x' 叫做 x 在 σ 下的象。

例 1 设函数 $y=f(x)$ 的定义域为 $[a,b]$，值域为 $[c,d]$，则 $f(\ )$ 即为由数集 $[a,b]$ 到数集 $[c,d]$ 的一个映射；

$\sigma(A)=\det A$，$A\in R^{n\times n}$，则 σ 是由 $R^{n\times n}$ 到实数集 R 的一个映射；

$$\sigma(f(x))\triangleq\frac{\mathrm{d}f(x)}{\mathrm{d}x}(=f'(x))$$

是由可导函数集

$$F[x]=\{f(x)\mid f'(x)\text{ 存在}\}$$

到导函数集

$$F'[x]=\left\{f'(x)\mid f'(x)=\frac{\mathrm{d}f(x)}{\mathrm{d}x},f(x)\in F[x]\right\}$$

的一个映射。

定义 2 设 $V(F)$ 与 $W(F)$ 是两个线性空间，σ 是由 $V(F)$ 到 $W(F)$ 的一个映射，且

$$\sigma(x+y)=\sigma(x)+\sigma(y)\qquad\forall\, x,y\in V(F)$$

$$\sigma(\lambda x)=\lambda\sigma(x)\qquad\forall\,\lambda\in F,\forall\, x\in V(F)$$

则称映射 σ 为由 $V(F)$ 到 $W(F)$ 的**线性映射**，特别，若还有 $W(F)\subset V(F)$，则称线性映射 σ 为 $V(F)$ 上的**线性变换**。

在例 1 中，$\sigma(A)=\det A$ 不是线性映射；$\sigma(f(x))=f'(x)$ 是线性映射，如果将 $F[x]$ 换为 $P[x]$，则 $\sigma(f(x))$ 是 $P[x]$ 上的线性变换；当 $A\in R^{n\times n}$，$x\in R^n$ 时，Ax 是 R^n 的线性变换；恒等变换 $\sigma(x)=x$ 是线性变换；零变换 $O(x)=\theta$ 是线性变换。

线性映射有以下性质：

定理 1 设 $\sigma:V(F)\to U(F)$ 是线性映射，则

a) $\sigma(\theta)=\theta'$，其中 θ 与 θ' 分别为 $V(F)$ 与 $U(F)$ 的零元素；

b) $\sigma(-x)=-\sigma(x)\qquad\forall\, x\in V(F)$；

c) σ 将 $V(F)$ 中的线性相关组映射为 $U(F)$ 中的线性相关组；

d) 设 $V_1(F)$ 是 $V(F)$ 的子空间，若

$$\sigma(V_1(F))=\{\sigma(x)\mid x\in V_1(F)\}^{①}$$

则 $\sigma(V_1(F))$ 是 $U(F)$ 的子空间，且

---

① $\sigma(V(F))$ 称为 σ 在 $V(F)$ 上的值域或 $V(F)$ 在 σ 下的象空间，记为 $\mathrm{Ima}(\sigma)$。

$$\dim\sigma(V_1(F)) \leqslant \dim V_1(F)$$

定理 2 设有线性映射 $\sigma_i: V(F) \to U(F)(i \in \underline{m})$ 及 μ: $U(F) \to W(F)$，则

$$\sigma = \sum_{i=1}^{m} \lambda_i \sigma_i \qquad \lambda_i \in F, i \in \underline{m}$$

是由 $V(F)$ 到 $U(F)$ 的线性映射，而

$$\mu\sigma_i \qquad i \in \underline{m}$$

是由 $V(F)$ 到 $W(F)$ 的线性映射。

定理 3 设 $\sigma: V(F) \to U(F)$ 为线性映射，$W(F)$ 与 $W_1(F)$ 是 $V(F)$ 的子空间，则

a) $\sigma(W(F) + W_1(F)) = \sigma(W(F)) + \sigma(W_1(F))$

b) $\sigma(W(F) \cap W_1(F)) \subset \sigma(W(F)) \cap \sigma(W_1(F))$

以上三个定理的证明略。

例 2 设 $V(F) = U(F) = R^2 \quad (F = R)$，且

$$\sigma = A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} : R^2 \to R^2$$

$$W(F) = \mathrm{span}\left[\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right], W_1(F) = \mathrm{span}\left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right]$$

是 R^2 的两个子空间，由于

$$W(F) \cap W_1(F) = \{\theta\}$$

故

$$A(W(F) \cap W_1(F)) = A(\theta) = \{\theta\}$$

而

$$A(W(F)) = \mathrm{span}\left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right] = A(W_1(F))$$

从而可知

$$A(W(F) \cap W_1(F)) \neq A(W(F)) \cap A(W_1(F))$$

定理 4 设 $\sigma: V_n(F)\to V_n(F)$ 是线性变换，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 为 $V_n(F)$ 的基，则

a) 若 $\sigma_1: V_n(F)\to V_n(F)$ 为另一线性变换，且 $\sigma_1(\varepsilon_i)=\sigma(\varepsilon_i), i\in \underline{n}$，则 $\sigma_1=\sigma$；

b) 存在唯一的 $A\in F^{n\times n}$，使

$$\sigma(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n) = (\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)A$$

并称 A 为**线性变换 σ 在基 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的矩阵**；

c) 对任意 $x\in V_n(F)$，若有 $a,b\in F^n$，使

$$x = (\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)a$$

$$\sigma(x) = (\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)b$$

则

$$b = Aa$$

d) ①$\sigma(V_n(F))=\text{span}[\sigma(\varepsilon_1),\sigma(\varepsilon_2),\cdots,\sigma(\varepsilon_n)]$

②$\dim[\sigma(V_n(F))]=\text{rank}A$

③$\dim(\sigma(V_n(F)))+\dim\text{Ker}(\sigma)=n$

其中 $\text{Ker}(\sigma)=\{x\mid\sigma(x)=\theta, x\in V_n(F)\}\subset V_n(F)$ 称为 $V_n(F)$ 在线性映射 σ 下的核或化零空间。

证 *a*)设任意 $x\in V_n(F)$，且

$$x = \alpha_1\varepsilon_1 + \alpha_2\varepsilon_2 + \cdots + \alpha_n\varepsilon_n$$

则由

$$\begin{aligned}\sigma(x) &= \alpha_1\sigma(\varepsilon_1) + \alpha_2\sigma(\varepsilon_2) + \cdots + \alpha_n\sigma(\varepsilon_n)\\ &= \alpha_1\sigma_1(\varepsilon_1) + \alpha_2\sigma_1(\varepsilon_2) + \cdots + \alpha_n\sigma_1(\varepsilon_n)\\ &= \sigma_1(x)\end{aligned}$$

知

$$\sigma = \sigma_1$$

b)设 $\sigma(\varepsilon_i)=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)a_i, a_i\in F^n, i\in\underline{n}$，取

$$A = (a_1, a_2, \cdots, a_n)$$

则有

$$\sigma(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n) = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)A$$

又由坐标的唯一性知 A 唯一；

c)因为

$$\begin{aligned}\sigma(x) &= \sigma[(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)a] \\ &= [\sigma(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)]a \\ &= (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)Aa \\ &= (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)b\end{aligned}$$

由坐标的唯一性有

$$b = Aa$$

d)①显然

②由①及§1.4定理1知

$$\dim[\sigma(V_n(F))] = \operatorname{rank}(\sigma(\varepsilon_1), \sigma(\varepsilon_2), \cdots, \sigma(\varepsilon_n))$$

又由 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 是 $V_n(F)$ 的基，则

$$\sigma(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n) = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)A$$

其中 A 由 $\sigma(\varepsilon_1), \sigma(\varepsilon_2), \cdots, \sigma(\varepsilon_n)$ 的坐标组成，于是(同构)

$$\operatorname{rank}[\sigma(\varepsilon_1), \sigma(\varepsilon_2), \cdots, \sigma(\varepsilon_n)] = \operatorname{rank}A$$

故

$$\dim[\sigma(V_n(F))] = \operatorname{rank}A$$

③设 $\dim[\operatorname{Ker}(\sigma)] = r$，$\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_r'$ 为 $\operatorname{Ker}(\sigma)$ 的基，由于 $\operatorname{Ker}(\sigma)$ 是 $V_n(F)$ 的子空间，故可由 $\operatorname{Ker}(\sigma)$ 的基 $\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_r'$ 扩充使 $\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_r', \varepsilon'_{r+1}, \cdots, \varepsilon_n'$ 为 $V_n(F)$ 的基。令

$$U(F) = \operatorname{span}[\varepsilon'_{r+1}, \varepsilon'_{r+2}, \cdots, \varepsilon_n']$$

于是

$$V_n(F) = \operatorname{Ker}(\sigma) \oplus U(F)$$

由于 $\sigma(\varepsilon_i')=\theta, i\in \underline{r}$ 及①，有

$$\sigma(V_n(F)) = \text{span}[\sigma(\varepsilon_1'),\sigma(\varepsilon_2'),\cdots,\sigma(\varepsilon_n')]$$

$$= \text{span}[\sigma(\varepsilon'_{r+1}),\sigma(\varepsilon'_{r+2}),\cdots,\sigma(\varepsilon_n')]$$

为证 $\sigma(\varepsilon'_{r+1}),\sigma(\varepsilon'_{r+2}),\cdots,\sigma(\varepsilon_n')$ 线性无关，令

$$\sum_{j=r+1}^{n} k_j\sigma(\varepsilon_j') = \theta \qquad k_j \in F, j = r+1,\cdots,n$$

有

$$\sigma(\sum_{j=r+1}^{n} k_j\varepsilon_j') = \theta$$

于是 $\sum\limits_{j=r+1}^{n} k_j\varepsilon_j' \in \text{Ker}(\sigma)$，故有

$$\sum_{j=r+1}^{n} k_j\varepsilon_j' = \sum_{i=1}^{r} h_i\varepsilon_i' \qquad h_i \in F, i \in \underline{r}$$

记 $-h_i=k_i \quad i\in \underline{r}$，则上式可变为

$$\sum_{i=1}^{n} k_i\varepsilon_i' = \theta$$

由于 $\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n'$ 是 $V_n(F)$ 的基，所以

$$k_i = 0 \qquad i \in \underline{n}$$

故 $\sigma(\varepsilon'_{r+1}),\sigma(\varepsilon'_{r+2}),\cdots,\sigma(\varepsilon_n')$ 线性无关，因此

$$\dim[\sigma(V_n(F))] = n - r$$

于是

$$\dim[\sigma(V_n(F))] + \dim\text{Ker}(\sigma) = n$$

推论 设 $A=(a_1,a_2,\cdots,a_n)\in F^{m\times n}$，$R(A)=\text{Span}[a_1,a_2,\cdots,a_n]\subset F^m$，$N(A)=\{x \mid Ax=\theta, x\in F^n\}$，则

$$\dim R(A) = \text{rank} A$$

$$\dim R(A) + \dim N(A) = n$$

注意：不能由

$$\dim[\sigma(V_n(F))] + \dim\mathrm{Ker}(\sigma) = n$$

就认为

$$V_n(F) = \sigma(V_n(F)) \oplus \mathrm{Ker}(\sigma)$$

例 3 在 n 维线性空间 $P[x]_n$ 中，定义线性变换

$$\sigma(f(x)) = \frac{\mathrm{d}f(x)}{\mathrm{d}x} \qquad f(x) \in P[x]_n$$

显然 $\sigma(P[x]_n)=P[x]_{n-1}$，$\mathrm{Ker}(\sigma)=P[x]_1$，但

$$\sigma(P[x]_n) + \mathrm{Ker}(\sigma) = P[x]_{n-1} \neq P[x]_n$$

例 4 若 $\sigma(A)=A^T$，$A\in R^{2\times 2}$，求 σ 在 $R^{2\times 2}$的基

$$\varepsilon_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \varepsilon_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \varepsilon_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \varepsilon_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

下的矩阵及 $\dim\sigma(R^{2\times 2})$。

解 因为

$$\sigma(\varepsilon_1) = \varepsilon_1{}^T = 1\varepsilon_1 + 0\varepsilon_2 + 0\varepsilon_3 + 0\varepsilon_4$$

$$= (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\sigma(\varepsilon_2) = \varepsilon_2{}^T = \varepsilon_3 = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\sigma(\varepsilon_3) = \varepsilon_3{}^T = \varepsilon_2 = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\sigma(\varepsilon_4) = \varepsilon_4^T = \varepsilon_4 = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

故有

$$\sigma(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

即 σ 在基 $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ 下对应的矩阵

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

显然 $\mathrm{rank}B=4$，于是由定理 4 的 $d)$②有

$$\dim\sigma(R^{2\times 2}) = \mathrm{rank}B = 4$$

例 5 设 $1, x, x^2, \cdots, x^{n-1}$ 为 $P[x]_n$ 的基，线性变换

$$\sigma(f(x)) = \frac{\mathrm{d}f(x)}{dx} \qquad f(x) \in P[x]_n$$

求 σ 在基 $1, x, x^2, \cdots, x^{n-1}$ 下的矩阵及 $\dim(\sigma(P[x]_n))$。

解 因为

$$\sigma(1) = 0 = (1, x, x^2, \cdots, x^{n-1})\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\sigma(x) = 1 = (1, x, x^2, \cdots, x^{n-1})\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\sigma(x^2) = 2x = (1, x, x^2, \cdots, x^{n-1})\begin{bmatrix} 0 \\ 2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

··

$$\sigma(x^{n-1}) = (n-1)x^{n-2} = (1, x, x^2, \cdots, x^{n-1})\begin{bmatrix} 0 \\ \vdots \\ 0 \\ n-1 \\ 0 \end{bmatrix}$$

于是

$$\sigma(1, x, x^2, \cdots, x^{n-1})$$

$$= (1, x, x^2, \cdots, x^{n-1})\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & 2 & \cdots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & n-1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

即 σ 在基 $1, x, x^2, \cdots, x^{n-1}$,下的矩阵为

$$A=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & 2 & \cdots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & n-1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

显然 $\operatorname{rank}A=n-1$.故

$$\dim\sigma(P[x]_n)=n-1$$

定理 5 设 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 与 $\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n'$ 为 $V_n(F)$ 的两个基.σ 为 $V_n(F)$ 的线性变换.且

$$\sigma(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)A$$

$$\sigma(\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n')=(\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n')B$$

$$(\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n')=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)T$$

则

$$B=T^{-1}AT$$

即 B 与 A 相似。

证 因为

$$\sigma(\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n')=(\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n')B$$
$$=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)TB$$

另一方面

$$\sigma(\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n')=\sigma[(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)T]$$
$$=[\sigma(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)]T=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)AT$$

于是

$$AT=TB$$

显然 T 可逆.故

$$B=T^{-1}AT$$

例 6 设

$$\varepsilon_1=\begin{bmatrix}1\\0\\2\\1\end{bmatrix},\varepsilon_2=\begin{bmatrix}0\\1\\0\\1\end{bmatrix},\varepsilon_3=\begin{bmatrix}0\\0\\1\\1\end{bmatrix},\varepsilon_4=\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

$$\varepsilon_1'=\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\varepsilon_2'=\begin{bmatrix}1\\1\\3\\0\end{bmatrix},\varepsilon_3'=\begin{bmatrix}1\\0\\1\\0\end{bmatrix},\varepsilon_4'=\begin{bmatrix}1\\0\\0\\1\end{bmatrix}$$

为 R^4 的两个基，且线性变换 σ 有

$$\sigma(\varepsilon_1')=\begin{bmatrix}1\\1\\0\\0\end{bmatrix},\sigma(\varepsilon_2')=\begin{bmatrix}1\\1\\1\\0\end{bmatrix}$$

$$\sigma(\varepsilon_3')=\begin{bmatrix}1\\1\\1\\1\end{bmatrix},\sigma(\varepsilon_4')=\begin{bmatrix}1\\0\\1\\1\end{bmatrix}$$

求 σ 在基 $\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4$ 下的矩阵。

解 设

$$(\varepsilon_1',\varepsilon_2',\varepsilon_3',\varepsilon_4')=(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)T$$

将两个基代入得

$$T=\begin{bmatrix}1&1&1&1\\0&1&0&0\\-2&1&-1&-2\\1&-3&0&2\end{bmatrix}$$

设

$$\sigma(\varepsilon_1',\varepsilon_2',\varepsilon_3',\varepsilon_4')=(\varepsilon_1',\varepsilon_2',\varepsilon_3',\varepsilon_4')B$$

将$(\varepsilon_1', \varepsilon_2', \varepsilon_3', \varepsilon_4')$与$\sigma(\varepsilon_1'), \sigma(\varepsilon_2'), \sigma(\varepsilon_3'), \sigma(\varepsilon_4')$代入得

$$B = \begin{bmatrix} 3 & 2 & 1 & -1 \\ 1 & 1 & 1 & 0 \\ -3 & -2 & -2 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

设

$$\sigma(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)A$$

由定理 5 有

$$A = TBT^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 1 & -1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

§1.6 线性空间的同构

设$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$是$V_n(F)$的基，$x \in V_n(F)$有

$$x = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

于是$x \in V_n(F)$与$a = (\alpha_1, \alpha_2, \cdots, \alpha_n)^T \in F^n$一一对应。

又若$y = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)b = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \in V_n(F)$，则

$x + y \in V_n(F)$对应$a + b \in F^n$，$\lambda x \in V_n(F)$对应$\lambda a \in F^n$，$\lambda \in F$，因此有可能将$V_n(F)$的问题转化为F^n的相应问题来讨论，这就是同构的概念。

定义 设 σ 是由 $V_n(F)$ 到 $U_m(F)$ 的线性映射，且

a) $\sigma[V_n(F)]=U_m(F)$

b) 若 $x_1,x_2\in V_n(F)$，$\sigma(x_1)=\sigma(x_2)$，则

$$x_1 = x_2$$

那么称 σ 为 $V_n(F)$ 与 $U_m(F)$ 间的一个**同构映射**。

若 $V_n(F)$ 与 $U_m(F)$ 之间可以建立一个同构映射，则称 $V_n(F)$ 与 $U_m(F)$ 的**同构的线性空间**，简称 $V_n(F)$ 与 $U_m(F)$ **同构**。

例 1 $V_n(F)$ 与 F^n 同构。

解 设 $(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)$ 是 $V_n(F)$ 的基，$x=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)a$，$a\in F^n$，显然

$$\sigma(x) = a$$

σ 是由 $V_n(F)$ 到 F^n 的线性映射，而且是同构映射，故 $V_n(F)$ 与 F^n 同构。

此例表明，即使 $V_n(F)$ 可以代表不同的线性空间，其元素可能很不相同，但利用同构关系，都可将 $V_n(F)$ 的问题转化到 F^n 的相应问题来讨论，可以说在同构意义下，$V_n(F)$ 与 F^n 的相应问题在本质上是一样的，而 F^n 的问题是具体的。上节定理 4)a)②的证明，正是利用了 $V_n(F)$ 与 F^n 同构这一事实。

定理 线性空间 $V_n(F)$ 与 $U_m(F)$ 同构的充要条件为 $n=m$。

证 必要性：设 $V_n(F)$ 与 $U_m(F)$ 同构，则在 $V_n(F)$ 与 $U_m(F)$ 间存在同构映射 σ，若 $(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)$ 为 $V_n(F)$ 的基，易知

$$U_m(F) = \text{span}[\sigma(\varepsilon_1),\sigma(\varepsilon_2),\cdots,\sigma(\varepsilon_n)]$$

且若

$$(\sigma(\varepsilon_1),\sigma(\varepsilon_2),\cdots,\sigma(\varepsilon_n))a = \theta_U \quad a \in F^n$$

其中 θ_U 表示 $U_m(F)$ 的零元素，即

$$\sigma[(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)a] = \theta_U$$

由于 σ 是同构映射，故必有

$$(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)a = \theta_V$$

其中 θ_V 是 $V_n(F)$ 的零元素，故必有 $a=\theta$（θ 为 F^n 的零元素），即 $\sigma(\varepsilon_1),\sigma(\varepsilon_2),\cdots,\sigma(\varepsilon_n)$ 线性无关，于是

$$\dim U_m(F) = n(=m)$$

充分性：设 $n=m$，且 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 与 $\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n'$ 分别为 $V_n(F)$ 与 $U_n(F)$ 的基，令

$$x = (\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\tilde{x} \qquad \tilde{x} \in F^n$$

$$\sigma(x) = (\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n')\tilde{x}$$

易证 σ 是由 $V_n(F)$ 到 $U_n(F)$ 的同构映射，故 $V_n(F)$ 与 $U_n(F)$ 同构。

例 2 设 $A\in R^{m\times n}$，则存在 $S\subset R^n$ 与 $R(A)$ 同构，其中

$$R(A) = \text{span}[a_1,a_2,\cdots,a_n]$$

$$A = (a_1,a_2,\cdots,a_n) \quad a_i \in R^m, i \in \underline{n}$$

证 设 $\text{rank}A=r$，则由上节定理 4 的推论有

$$\dim R(A) = r$$

$$\dim\text{Ker}(\sigma) = \dim N(A) = n - r$$

其中 $\sigma(x)=Ax, x\in R^n$。于是存在 R^n 的子空间 S，使

$$R^n = S \oplus N(A)$$

由于 $\dim S=r$，故 S 与 $R(A)$ 同构。

§1.7 不变子空间

在 $V(F)$ 的子空间中，不变子空间是一类特别有用的子空间，下面作简要介绍。

定义 设 σ 是线性空间 $V(F)$ 的线性变换，且 $U(F)$ 是

$V(F)$的子空间，若 $\sigma[U(F)]\subset U(F)$，则称 $U(F)$ 为 $V(F)$ 的关于 σ 的**不变子空间**。

显然，$V_n(F)$的平凡子空间是 $V_n(F)$的任一线性变换的不变子空间；$P[x]_m$ 是 $P[x]_n(m\leqslant n)$关于导数变换的不变子空间。

例 设 σ 是 $V_n(F)$的线性变换，若 $x\in V_n(F)$时，有 $\sigma^{m-1}(x)\neq\theta,\sigma^m(x)=\theta$，则

$$S=\text{span}[x,\sigma(x),\sigma^2(x),\cdots,\sigma^{m-1}(x)]\subset V_n(F)$$

是 σ 的不变子空间，并称 S 为由 x 经 σ 生成的**循环子空间**。

事实上，$x,\sigma(x),\sigma^2(x),\cdots,\sigma^{m-1}(x)$是线性无关的（请读者自己证明），故它是 S 的基，若 $y\in S$，则有 $\alpha\in F^m$，使

$$y=(x,\sigma(x),\sigma^2(x),\cdots,\sigma^{m-1}(x))\alpha$$

于是

$$\begin{aligned}\sigma(y)&=\sigma[(x,\sigma(x),\sigma^2(x),\cdots,\sigma^{m-1}(x))\alpha]\\&=(\sigma(x),\sigma^2(x),\cdots,\sigma^{m-1}(x),\theta)\quad\alpha\in S\end{aligned}$$

即 S 是 $V_n(F)$关于 σ 的不变子空间。

定理 设 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_r$ 是 $V_n(F)$的子空间 $U_r(F)$的基，经扩充后，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_r,\varepsilon_{r+1},\cdots,\varepsilon_n$ 为 $V_n(F)$的基，则 $U_r(F)$是 $V_n(F)$的线性变换 σ 的不变子空间的充要条件为 σ 在基 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的矩阵 A，即

$$\sigma(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)A$$

中的 A 为上三角块矩阵

$$A=\begin{pmatrix}A_1 & A_2\\ O & A_3\end{pmatrix}$$

其中 $A_1\in F^{r\times r},A_2\in F^{r\times(n-r)},A_3\in F^{(n-r)\times(n-r)},O=O_{(n-r)\times r}$。

证明留给读者自己练习。

推论 设 $V_i\subset V_n(F),i\in\underline{m}$是关于线性变换 σ 的不变子

空间，$\varepsilon_1^{(i)},\varepsilon_2^{(i)},\cdots,\varepsilon_{a_i}^{(i)}$为 V_i 的基，$i\in \underline{m}$，$\sum\limits_{i=1}^{m}a_i=n$，且

$$V_n(F)=\bigoplus_{i=1}^{m}V_i$$

则 σ 在 $V_n(F)$的基 $\varepsilon_1^{(1)},\varepsilon_2^{(1)},\cdots,\varepsilon_{a_1}^{(1)},\varepsilon_1^{(2)},\varepsilon_2^{(2)},\cdots,\varepsilon_{a_2}^{(2)},\cdots,\varepsilon_1^{(m)},\varepsilon_2^{(m)},\cdots,\varepsilon_{a_m}^{(m)}$下的矩阵 A 为对角块矩阵，即

$$A=\begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_m \end{bmatrix}$$

其中 $A_i\in F^{a_i\times a_i}$，$i\in \underline{m}$。

由此可见，若能将 $V_n(F)$分解为不变子空间的直和将是十分方便的。

习 题 一

一、验证下列集合，在指定的数域和运算下是否为线性空间？

1. 设 $V=\left\{x\,|\,x=\begin{pmatrix}x_1\\x_2\end{pmatrix},x_1,x_2\in R\right\}$，若 $x=\begin{pmatrix}x_1\\x_2\end{pmatrix}$，$y=\begin{pmatrix}y_1\\y_2\end{pmatrix}$，$x+y\triangleq(x_1+y_1,x_2+y_2+x_1y_1)^T$，$\lambda x\triangleq(\lambda x_1,\lambda x_2+\dfrac{\lambda(\lambda-1)}{2}x_1^2)^T$

2. 给定 $A_0\in F^{n\times n}$，V 是所有满足 $A_0B=BA_0$ 的 B 的集合，关于矩阵的加法运算、数与矩阵的乘积运算。

二、1. 证明

$$S=\left\{A\,|\,A=\begin{pmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{pmatrix}\in R^{2\times 2},a_{11}+a_{22}=0\right\}$$

是 $R^{2\times3}$ 的线性子空间，并求其维数。

2. 证明

$$S = \{A \mid A \in R^{n\times n} \text{ 且 } A = A^T\}$$

是 $R^{n\times n}$ 的线性子空间，并求其基。

3. 求由

$$\alpha_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 1 \end{bmatrix}, \alpha_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \alpha_3 = \begin{bmatrix} -1 \\ 1 \\ -3 \\ 0 \end{bmatrix}, \alpha_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

所生成的线性空间 $\text{span}\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ 的基。

三、在 R^4 中，求 $a=(1,2,1,1)^T$ 在基

$$e_1' = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, e_2' = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix},$$

$$e_3' = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, e_4' = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$$

下的坐标。

四、设

$$e_1' = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}, e_2' = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix},$$

$$e_3' = \begin{pmatrix} -1 \\ 2 \\ 1 \\ 1 \end{pmatrix}, e_4' = \begin{pmatrix} -1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

及

$$e_1'' = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}, e_2'' = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 2 \end{pmatrix},$$

$$e_3'' = \begin{pmatrix} -2 \\ 1 \\ 1 \\ 2 \end{pmatrix}, e_4'' = \begin{pmatrix} 1 \\ 3 \\ 1 \\ 2 \end{pmatrix}$$

为 R^4 的两个基，求基的过渡矩阵及 $\alpha=(x_1,x_2,x_3,x_4)^T$ 在基 $\{e_1',e_2',e_3',e_4'\}$ 下的坐标。

五、设 $\alpha,\beta,\gamma\in V(F)$，有 $k_i\in F$，$i\in \underline{3}$，使

$$k_1\alpha+k_2\beta+k_3\gamma=\theta \qquad k_1k_3\neq 0$$

求证

$$\text{span}\{\beta,\gamma\}=\text{span}\{\alpha,\beta\}$$

六、设 $U(F)$、$W(F)$ 为线性空间 $V(F)$ 的子空间，$\alpha\in U(F)$，$\beta\in W(F)$。求证

1. 当 $\alpha,\beta\notin U(F)\cap W(F)$ 时，α、β 线性无关；

2. 若另有 $\gamma\in U(F)\cap W(F)$，但 $\gamma\neq\theta$，则 α、β、γ 线性无关。

七、设 $U(F)$、$W(F)$ 为线性空间 $V(F)$ 的两个真子空间，$\alpha\in U(F)$，$\alpha\notin W(F)$，β 为 $V(F)$ 中任一元素，求证

1. 在所有形如 $\alpha+\lambda\beta(\lambda\in F)$ 的元素中，最多只有一个元

素落在 $W(F)$ 中；

2. 当 $\beta\in U(F)$，$\lambda\neq 0$ 时，所有的线性组合 $\alpha+\lambda\beta$ 均不落在 $U(F)$ 中。

八、求证：有限个真子空间覆盖不了整个线性空间，即当 V_1、V_2、…、V_s 均为 $V(F)$ 的真子空间时，则 $V(F)$ 中至少有一个元素不属于 V_1、V_2、…、V_s 中的任何一个。

九、判断下面定义的映射是否为线性变换：

1. 设 $\alpha_0\in V(F)$ 为一固定元素，定义映射

$$\sigma(\alpha)=\alpha+\alpha_0 \qquad \forall\ \alpha\in V(F)$$

2. 在 R^3 中定义映射

$$\sigma(x)=\begin{bmatrix} x_1^2 \\ x_1+x_2 \\ x_3^2 \end{bmatrix},x=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\in R^3$$

十、证明：

1. $R(A+B)\subset R(A)+R(B) \quad \forall\ A,B\in R^{m\times n}$

2. $\text{rank}(A+B)\leqslant\text{rank}A+\text{rank}B \quad \forall\ A,B\in R^{m\times n}$

3. 若 $A,B\in C^{n\times n}$ 且 $AB=O$，则

$$\text{rank}A+\text{rank}B\leqslant n$$

4. 若 $A\in C^{n\times n}$，$A^2=I_n$，则

$$\text{rank}(A+I_n)+\text{rank}(A-I_n)=n$$

5. 设 $A,B\in C^{m\times n}$，证明

$$N(A)\cap N(B)=N(A+B)\cap N(A-B)$$

6. 设 $A,B\in C^{m\times n}$，证明

$$|\text{rank}A-\text{rank}B|\leqslant\text{rank}(A+B)$$

7. 设 $A\in F^{m\times n}$，$B\in F^{n\times p}$，证明

$$\text{rank}AB\leqslant\min(\text{rank}A,\text{rank}B)$$

$$\text{rank}A+\text{rank}B-n\leqslant\text{rank}AB$$

十一、设 σ 为 R^3 的线性变换

$$e_1=\begin{bmatrix}1\\0\\0\end{bmatrix},e_2=\begin{bmatrix}0\\1\\0\end{bmatrix},e_3=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$e_1'=\begin{bmatrix}-1\\0\\2\end{bmatrix},e_2'=\begin{bmatrix}0\\1\\1\end{bmatrix},e_3'=\begin{bmatrix}3\\-1\\0\end{bmatrix}$$

为 R^3 的两个基，且

$$\sigma(e_1')=\begin{bmatrix}-5\\0\\3\end{bmatrix},\sigma(e_2')=\begin{bmatrix}0\\-1\\6\end{bmatrix},$$

$$\sigma(e_3')=\begin{bmatrix}-5\\-1\\9\end{bmatrix}$$

求 σ 在基 e_1,e_2,e_3 下对应的矩阵。

十二、设 $\alpha_1,\alpha_2,\cdots,\alpha_n;\beta_1,\beta_2,\cdots,\beta_n;\gamma_1,\gamma_2,\cdots,\gamma_n$ 为 $V_n(F)$的三个基，且

$$[\beta_1,\beta_2,\cdots,\beta_n]=[\alpha_1,\alpha_2,\cdots,\alpha_n]T_1$$
$$T_1\in F^{n\times n}$$
$$[\gamma_1,\gamma_2,\cdots,\gamma_n]=[\alpha_1,\alpha_2,\cdots,\alpha_n]T_2$$
$$T_2\in F^{n\times n}$$

σ 为 $V_n(F)$的线性变换，且

$$\sigma[\gamma_1,\gamma_2,\cdots,\gamma_n]=[\alpha_1,\alpha_2,\cdots,\alpha_n]A$$
$$A\in F^{n\times n}$$

求 σ 在基 $\beta_1,\beta_2,\cdots,\beta_n$ 下对应的矩阵。

十三、设 σ 是 $V_n(F)$的线性变换，若 $\sigma^{n-1}(x)\neq\theta,\sigma(x)=\theta,x\in V_n(F)$，求 $V_n(F)$的基及 σ 在该基下对应的矩阵。

十四、设 $A\in C^{n\times n}$，$B\in C^{m\times m}$，则由

$$Y = AX + XB$$

定义 $\sigma(X)=Y$ 是 $C^{n\times m}$ 上的线性变换。

十五、设 σ 是由 $V_n(F)$ 到 $U_m(F)$ 的线性映射，$\alpha_1,\alpha_2,\cdots,\alpha_n$，$\alpha_1',\alpha_2',\cdots,\alpha_n'$ 为 $V_n(F)$ 的两组基，$\beta_1,\beta_2,\cdots,\beta_m$，$\beta_1',\beta_2',\cdots,\beta_m'$ 为 $U_m(F)$ 的两组基，且

$$\sigma[\alpha_1,\alpha_2,\cdots,\alpha_n] = [\beta_1,\beta_2,\cdots,\beta_m]A$$

$$\sigma[\alpha_1',\alpha_2',\cdots,\alpha_n'] = [\beta_1',\beta_2',\cdots,\beta_m']B$$

问矩阵 A 和 B 之间有何关系？

十六、证明特征子空间的和是直和。

第二章　酉空间与欧氏空间

在 R^3 中，向量的长度及两个向量的夹角的度量是用向量的"内积"来刻画的，这里将内积的概念推广到一般的线性空间。在已经定义内积的线性空间中，酉空间和欧氏空间是最常用的。

§2.1　酉空间、欧氏空间

定义 1　设 A、B 为两个集合

$$A\times B=\{(a,b)\mid a\in A,b\in B\}$$

称为集合 A 与集合 B 的**直积**。

例 1　设 $A=[1,2]$，$B=[2,4]$，则

$$A\times B=\{(x,y)\mid 1\leqslant x\leqslant 2,2\leqslant y\leqslant 4\}$$

在 xy 坐标面上代表一个矩形。

定义 2　在线性空间 $V(C)$ 上，若映射 (x,y)：

$$V(C)\times V(C)\to C$$

满足

a)　$(x,y)=\overline{(y,x)}$　　　$\forall\ x,y\in V(C)$

b)　$(\lambda x,y)=\overline{\lambda}(x,y)$　　　$\forall\ \lambda\in C,\forall\ x,y\in V(C)$

c)　$(x+y,z)=(x,z)+(y,z)$　$\forall\ x,y,z\in V(C)$

则称 (x,y) 为定义在 $V(C)$ 上的**内积**；定义了内积的线性空间称为**内积空间**；

d)　$(x,x)\geqslant 0,\forall\ x\in V(C)$，当且仅当 $x=\theta$ 时等号成立；

e) $(x,x)\geqslant 0,\forall\ x\in V(C)$

若内积(x,y)满足 a)—d)，则称内积(x,y)**具正定度量**；若内积(x,y)满足 a)、b)、c)、e)，则称内积(x,y)**具半正定度量或非负度量**。

在线性空间 $V_n(C)$ 中，定义了具正定度量的内积后，$V_n(C)$就称为**酉空间**，记为 $V_n(C、U)$。

例 2 若 $A=A^H\in C^{n\times n}$①，对任意的 $x,y\in C^n$，令

$$(x,y)=x^HAy \tag{1}$$

则它是 C^n 上的内积。

解 由(1)式知，$(x,y)=x^HAy$ 是由 $C^n\times C^n\to C$ 的映射。设 $x=(x_1,x_2,\cdots,x_n)^T$，$y=(y_1,y_2,\cdots,y_n)^T$，$A=(a_{ij})_{n\times n}$，则

$$(x,y)=\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{x_i}a_{ij}y_j=\overline{\overline{\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{\overline{x_i}a_{ij}y_j}}}$$

$$=\overline{\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{y_j}a_{ij}x_i}=\overline{y^HAx}=\overline{(y,x)}$$

$$(\lambda x,y)=(\lambda x)^HAy=\bar{\lambda}x^HAy=\bar{\lambda}(x,y)\quad\forall\ \lambda\in C$$

$$(x+y,z)=(x+y)^HAz=(x^H+y^H)Az$$

$$=x^HAz+y^HAz=(x,z)+(y,z)$$

于是由内积的定义知，$(x,y)=x^HAy$ 是 C^n 上的内积。特别，当 $A=I_n$ 时，$(x,y)=x^Hy$ 称为标准内积，今后如不特殊声明，内积均采用标准内积。

例 3 设 $A,B\in C^{n\times n}$则

$$(A,B)=tr(A^HB)$$

是 $C^{n\times n}$上的内积。

有了内积概念后，可以仿照 R^3 中用内积来定义$V_n(C,U)$

---

① $A^H\triangleq(\overline{A})^T$，若 $A=A^H$，则称 A 为埃尔米特(Hermite)矩阵。

中向量的长度。

定义 3 对任意的 $x\in V_n(C,U)$称

$$\|x\| = (x,x)^{\frac{1}{2}} \tag{2}$$

为向量 x 的长度或模。

下面讨论内积的性质。

定理 1 设(x,y)为 $V(C)$上的内积,则

$$\left(\sum_{i=1}^{m}\lambda_i x_i,\sum_{j=1}^{r}\mu_j y_j\right)=\sum_{i=1}^{m}\sum_{j=1}^{r}\overline{\lambda}_i\mu_j(x_i,y_j)$$

$$\forall\ \lambda_i,\mu_j\in C\quad \forall\ x_i,y_j\in V(C)\quad i\in \underline{m},j\in \underline{r}$$

这由内积定义可以直接导出。

定理 2 若(x,y)为 $V_n(C)$上具正定度量的内积,则

a) $|(x,y)|\leqslant\|x\|\ \|y\| \qquad \in x,y\in V_n(C)$ (3)

不等式(3)称为柯西-希瓦尔兹(Cauchy-Schwarz)不等式;

b) $\|x+y\|\leqslant\|x\|+\|y\| \qquad \forall\ x,y\in V_n(C)$ (4)

且 a)、b)中等号成立的充要条件为 x 与 y 线性相关。

证 a) 由于(x,y)是具正定度量的内积,则对任意的 $\theta\neq y\in V_n(C)$有

$$0\leqslant\left(x-\frac{\overline{(x,y)}}{(y,y)}y,x-\frac{\overline{(x,y)}}{(y,y)}y\right)$$

$$=(x,x)-2\frac{|(x,y)|^2}{(y,y)}+\frac{|(x,y)|^2}{(y,y)^2}(y,y)$$

$$=(x,x)-\frac{|(x,y)|^2}{(y,y)}$$

从而有

$$(x,x)(y,y)-|(x,y)|^2\geqslant 0$$

即不等式(3)成立。(3)的等号成立等价于

$$x=\frac{(x,y)}{(y,y)}y$$

即 x 与 y 线性相关。

当 $y=\theta$ 时，(3)显然成为等式，且此时 x 与 y 是线性相关的。

b) 由 a)有

$$\begin{aligned}\|x+y\|^2 &= (x+y,x+y) = (x,x) + 2\mathrm{Re}(x,y) \\ &+ (y,y) \leqslant (x,x) + 2|(x,y)| + (y,y) \leqslant \|x\|^2 \\ &+ 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2\end{aligned}$$

故不等式(4)成立。

定义 4 设 $e_1,e_2,\cdots,e_n$ 为 $V_n(C,U)$的基，(x,y)为其内积，若

$$a_{ij} = (e_i,e_j) \qquad i,j \in \underline{n}$$

则称 $A=(a_{ij})\in C^{n\times n}$为内积$(x,y)$在基 $e_1,e_2,\cdots,e_n$ 下的矩阵。

例 4 在 C^n 中取基

$$e_i = (0,0,\cdots,0,\overbrace{1}^{\text{第 } i \text{ 个分量}},0,\cdots,0)^T \qquad i \in \underline{n}$$

当 $x=(x_1,x_2,\cdots,x_n)^T$, $y=(y_1,y_2,\cdots,y_n)^T\in C^n$ 时，定义 $(x,y)=\sum_{i=1}^{n}\bar{x}_i y_i$(具正定度量)，且它在这个基下对应的矩阵为单位阵 I_n。

定理 3 设 $A\in C^{n\times n}$为 $V_n(C)$中内积(x,y)在基 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的矩阵，则

a) $A=A^H$

b) $(x,y)=\tilde{x}^H A\tilde{y}$

其中 $x=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\tilde{x}$, $y=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\tilde{y}$, $\tilde{x}=(x_1,x_2,\cdots,x_n)^T$, $\tilde{y}=(y_1,y_2,\cdots,y_n)^T\in C^n$。

c) (x,y)为具正定度量的内积的充要条件为，对任意 $\theta\neq\tilde{x}\in C^n$，均有

$$\tilde{x}^H A \tilde{x} > 0$$

证 a)因为

$$a_{ij} = (\varepsilon_i, \varepsilon_j) = \overline{(\varepsilon_i, \varepsilon_j)} = \bar{a}_{ji} \qquad i, j \in \underline{n}$$

故

$$A = A^H \qquad A = (a_{ij}) \in C^{n\times n}$$

b)

$$(x, y) = \left(\sum_{i=1}^{n} x_i \varepsilon_i, \sum_{j=1}^{n} y_j \varepsilon_j\right)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \bar{x}_i (\varepsilon_i, \varepsilon_j) y_j$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \bar{x}_i a_{ij} y_j = \tilde{x}^H A \tilde{y};$$

c) 显然。

定理 4 设内积(x,y)在$V_n(C)$的两个基$[\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]$与$[\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_n']$下的矩阵分别为$A$与$B$，且

$$[\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_n'] = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]T$$

则

$$B = T^H A T$$

此时称B与A合同。

证 设$A = (a_{ij})_{n\times n}$，$B = (b_{ij})_{n\times n}$，$T = (t_{ij})_{n\times n}$，由于

$$b_{ij} = (\varepsilon_i', \varepsilon_j') = \left(\sum_{k=1}^{n} t_{ki}\varepsilon_k, \sum_{l=1}^{n} t_{lj}\varepsilon_l\right)$$
$$= \sum_{k=1}^{n}\sum_{l=1}^{n} \bar{t}_{ki}(\varepsilon_k, \varepsilon_l) t_{lj} = \sum_{k=1}^{n}\sum_{l=1}^{n} \bar{t}_{ki} a_{ij} t_{lj}$$

故

$$B = T^H A T$$

例 5 设

$$\varepsilon_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \varepsilon_2 = \begin{bmatrix} 0 \\ i \end{bmatrix}$$

$$\varepsilon_1' = \begin{bmatrix} 1 \\ i \end{bmatrix} \qquad \varepsilon_2' = \begin{bmatrix} 1 \\ 2i \end{bmatrix}$$

(其中 $i=\sqrt{-1}$)为 C^2 的两个基，内积

$$(x,y) = x^H D y \qquad \forall\ x,y \in C^2$$

其中

$$D = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$$

求 (x,y) 在这两个基下的矩阵。

解 $(\varepsilon_1,\varepsilon_1)=\varepsilon_1^H D\varepsilon_1=(1,0)\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}=1$，同理可得 $(\varepsilon_1,\varepsilon_2)=-1$，$(\varepsilon_2,\varepsilon_1)=-1$，$(\varepsilon_2,\varepsilon_2)=1$，于是 (x,y) 在基 $\varepsilon_1,\varepsilon_2$ 下的矩阵

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

又 $(\varepsilon_1',\varepsilon_1')=2$，$(\varepsilon_1',\varepsilon_2')=3+i$，$(\varepsilon_2',\varepsilon_1')=3-i$，$(\varepsilon_2',\varepsilon_2')=7$，于是 (x,y) 在基 $\varepsilon_1',\varepsilon_2'$ 下的矩阵

$$B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

如果先导出

$$(\varepsilon_1',\varepsilon_2') = (\varepsilon_1,\varepsilon_2)T$$

中的

$$T = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

也可由公式 $B=T^H AT$ 中求出 B。

在 $V(R)$ 上，可仿定义 2 定义内积如下：

定义 5 设映射 $(x,y):V(R)\times V(R)\to R$，且满足

a) $(x,y)=(y,x) \qquad \forall\ x,y\in V(R)$

b)　$(\lambda x,y)=\lambda(x,y)$　　$\forall\ \lambda\in R,\forall\ x,y\in V(R)$

c)　$(x+y,z)=(x,z)+(y,z)$　$\forall\ x,y,z\in V(R)$

则称(x,y)为$V(R)$上的内积。

d)　$(x,x)\geqslant 0$，当且仅当$x=\theta$时，等号成立，$x\in V(R)$

e)　$(x,x)\geqslant 0$　　　$\forall\ x\in V(R)$

当(x,y)满足a)～d)时，称(x,y)具**正定度量**，当(x,y)满足a)、b)、c)、e)时，称(x,y)具**半正定度量或非负度量**。

在线性空间$V_n(R)$中，定义了具正定度量的内积后，$V_n(R)$就称为**欧几里得(Euclid)空间**，简称为**欧氏空间**，并记为$V_n(R,E)$。

$$\|x\| = (x,x)^{\frac{1}{2}} \qquad \forall\ x\in V_n(R,E)$$

也称为向量x的**长度或模**。

例6　在R^n中，定义内积

$$(x,y)=\sum_{i=1}^{n}x_iy_i \qquad \forall\ x,y\in R^n$$

其中　$x=(x_1,x_2,\cdots,x_n)^T,y=(y_1,y_2,\cdots,y_n)^T$　$(x_i,y_i\in R,i\in\underline{n})$，则$R^n$成为$n$维欧氏空间。

与酉空间内积性质作相应的讨论，可得欧氏空间内积的相应性质，不再赘述。

§2.2　向量的正交与标准正交基

在R^3中，向量的垂直与直角坐标系的概念，推广到酉(欧氏)空间，即为向量的正交与标准正交基。显然，这也将为对酉(欧氏)空间的讨论提供方便。

定义1　若$x,y\in V_n(C,U)(V_n(R,E))$，且

$$(x,y)=0$$

则称向量 x 与向量 y 正交，记为 $x \perp y$。

例 1 在 R^2 中，取两种不同的具正定度量的内积

$$(x,y)_1 = x^T y \qquad \forall\, x,y \in R^2$$

$$(x,y)_2 = x^T A y \qquad \forall\, x,y \in R^2$$

其中

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

由此所得的两个欧氏空间，分别记为 $V_2(R,E_1)$ 与 $V_2(R,E_2)$，问向量 $x_0=(1,1)^T$、$y_0=(-1,1)^T$ 在这两个欧氏空间中是否正交？

解 由于

$$(x_0,y_0)_1 = (1,1)(-1,1)^T = 0$$

$$(x_0,y_0)_2 = (1,1)\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 1$$

故 x_0、y_0 在 $V_2(R,E_1)$ 中正交，在 $V_2(R,E_2)$ 中不正交。

此例表明，二向量正交与否，由该空间的内积所确定。

在酉(欧氏)空间中，向量的正交具有以下性质。

定理 1 在 $V_n(C,U)(V_n(R,E))$ 中，有下列性质：

$a)$ $(x,\theta)=0 \qquad \forall\, x \in V_n(C,U)(V_n(R,E))$

$b)$ $(x,y)=0$ 当且仅当 $(\lambda x,y)=0$，其中 $\lambda \in C(R)$，x，$y \in V_n(C,U)(V_n(R,E))$

$c)$ 不含 θ 的两两正交的向量组必线性无关，即若

$$\theta \neq x_i \in V_n(C,U)(V_n(R,E)) \quad i \in \underline{m}(m \leqslant n)$$

且

$$(x_i,x_j) = 0 \qquad i \neq j \quad i,j \in \underline{m}$$

则向量组 $[x_1,x_2,\cdots,x_m]$ 线性无关。

证 $a)$、$b)$ 显然；

c) 设 $a=(\alpha_1,\alpha_2,\cdots,\alpha_m)^T\in C^m(R^m)$，有

$$(x_1,x_2,\cdots,x_m)a=\theta$$

其中 $x_1,x_2,\cdots,x_m$ 两两正交。

由 a)有

$$0=(x_i,\theta)=(x_i,\sum_{k=1}^{m}\alpha_k x_k)$$

$$=\sum_{k=1}^{m}\alpha_k(x_i,x_k)=\alpha_i(x_i,x_i)\qquad i\in \underline{m}$$

由于 $x_i\neq\theta,i\in\underline{m}$ 且 $(x_i,x_i)>0$，故

$$\alpha_i=0\qquad i\in\underline{m}$$

所以正交向量组 $\{x_1,x_2,\cdots,x_m\}$ 线性无关。

定义 2 若 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 为 $V_n(C,U)(V_n(R,E))$ 的基，且

$$(\varepsilon_i,\varepsilon_j)=0\qquad i\neq j\quad i,j\in\underline{n}\tag{1}$$

则称 $\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n\}$ 为 $V_n(C,U)(V_n(R,E))$ 的**正交基**。若正交基 $\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n\}$ 还满足

$$(\varepsilon_i,\varepsilon_i)=1\qquad i\in\underline{n}\tag{2}$$

则称 $\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n\}$ 为 $V_n(C,U)(V_n(R,E))$ 的**标准正交基**。

对于标准正交基 $\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n\}$，(1)式与(2)式可合并写为

$$(\varepsilon_i,\varepsilon_j)=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}\qquad i,j\in\underline{n}$$

或

$$(\varepsilon_i,\varepsilon_j)=\delta_{ij}\qquad i,j\in\underline{n}$$

这里的 δ_{ij}，当 $i=j$ 时为 1，$i\neq j$ 时为零。

例 2 在 R^3 中定义内积

$$(x,y)=x^Ty\qquad \forall\, x,y\in R^3$$

易验证

$$e_1=(1,0,0)^T, e_2=(0,1,0)^T, e_3=(0,0,1)^T$$

$e_1'=\left(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2},0\right)^T$，$e_2'=\left(\frac{\sqrt{6}}{6},-\frac{\sqrt{6}}{6},\frac{\sqrt{6}}{3}\right)^T$，$e_3'=\left(-\frac{\sqrt{3}}{3},\frac{\sqrt{3}}{3},\frac{\sqrt{3}}{3}\right)$均为 $V_3(R,E)$ 的标准正交基。由此可见，标准正交基也不唯一。

例 3 若以 $P[x]_n$ 表示系数为实数，次数低于 n 次的多项式的全体所组成的 n 维线性空间，定义内积

$$(f,g)=\int_{-1}^{1}f(x)g(x)\mathrm{d}x \qquad \forall f(x),g(x)\in P[x]_n \tag{3}$$

则

$$\varepsilon_{i+1}=\sqrt{\frac{2i+1}{2}}L_i(x) \qquad i=0,1,2,\cdots,n-1$$

是 $P[x]_n$ 中的标准正交基，其中

$$L_i(x)=\frac{1}{2^i i!}\frac{\mathrm{d}^i}{\mathrm{d}x^i}(x^2-1)^i$$

称为勒让得(Legendre)多项式。

解

$$(\varepsilon_i,\varepsilon_j)=\int_{-1}^{1}\sqrt{\frac{2i-1}{2}}L_{i-1}(x)\sqrt{\frac{2j-1}{2}}L_{j-1}(x)\mathrm{d}x$$

$$=\frac{\sqrt{(2i-1)(2j-1)}}{2^{i+j-1}(i-1)!(j-1)!}\int_{-1}^{1}\frac{\mathrm{d}^{i-1}}{\mathrm{d}x^{i-1}}(x^2-1)^{i-1}$$

$$\times\frac{\mathrm{d}^{j-1}}{\mathrm{d}x^{j-1}}(x^2-1)^{j-1}\mathrm{d}x=\begin{cases}0 & i>j\\ 1 & i=j\end{cases}$$

由于积分号内对 i、j 是对称的，故 $j>i$ 时，也有 $(\varepsilon_i,\varepsilon_j)=0$，于是 $[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n]$ 在(3)式所定义的内积意义下，是 $P[x]_n$ 的标准正交基。

标准正交基具有下述性质。

定理 2 若 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 为 $V_n(C,U)(V_n(R,E))$ 的标准正交基，则

a) (x,y) 在该基下的矩阵为单位阵 I_n；

b) 若 $x=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\tilde{x}$，$y=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\tilde{y}$，其中 $\tilde{x}=(x_1,x_2,\cdots,x_n)^T$，$\tilde{y}=(y_1,y_2,\cdots,y_n)^T\in C^n(R^n)$ 有

$$(x,y)=\tilde{x}^H\tilde{y} \qquad (=\tilde{x}^T\tilde{y})$$

c) $x_i=(\varepsilon_i,x) \qquad i\in \underline{n}$

d) 若 $\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n'$ 也是 $V_n(C,U)(V_n(R,E))$ 的标准正交基，且

$$[\varepsilon_1',\varepsilon_2',\cdots,\varepsilon_n']=[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n]T$$

那么 T 是酉阵，即 $T^H=T^{-1}$（T 是正交阵，即 $T^T=T^{-1}$）。

我们只需证 $V_n(C,U)$ 的情形，对于 $V_n(R,E)$，由 $x^H=x^T$ 及前者可得。

证 a) 显然；

b) $$(x,y)=\left(\sum_{i=1}^n x_i\varepsilon_i,\sum_{j=1}^n y_j\varepsilon_j\right)$$

$$=\sum_{i=1}^n\sum_{j=1}^n \overline{x_i}(\varepsilon_i,\varepsilon_j)y_j=\sum_{i=1}^n \overline{x_i}y_i=\tilde{x}^H\tilde{y}$$

c) $$(\varepsilon_i,x)=\left(\varepsilon_i,\sum_{j=1}^n x_j\varepsilon_j\right)=x_i(\varepsilon_i,\varepsilon_i)=x_i,\ i\in\underline{n}$$

d) 设 $T=(t_1,t_2,\cdots,t_n)$，$t_i\in C^n$，$i\in\underline{n}$，由于 $\varepsilon_i'=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)t_i$，$i\in\underline{n}$，故由 b) 有

$$(\varepsilon_i',\varepsilon_j')=t_i^Ht_j=\delta_{ij}$$

于是

$$T^HT=I_n$$

即 $T^H=T^{-1}$。证毕。

对于酉（欧氏）空间，显然使用标准正交基特别方便，如何

将一般的基化为标准正交基？下面的定理给出了具体的方法。

定理 3 设$[\alpha_1,\alpha_2,\cdots,\alpha_m]$为$V_n(C,U)(V_n(R,E))$的线性无关向量组$(m\leqslant n)$，则存在正交向量组$[\beta_1,\beta_2,\cdots,\beta_m]\subset V_n(C,U)(V_n(R,E))$，使得

$$[\alpha_1,\alpha_2,\cdots,\alpha_m]=[\beta_1,\beta_2,\cdots,\beta_m]B$$

其中 $B\in C^{m\times m}(R^{m\times m})$为可逆上三角阵。

证 由于向量组$[\alpha_1,\alpha_2,\cdots,\alpha_m]$线性无关，故

$$\alpha_i\neq\theta \qquad i\in\underline{m}$$

取 $\beta_1=\alpha_1(\neq\theta)$，所以$(\beta_1,\beta_1)>0$。令

$$\beta_2=\alpha_2-b_{12}\beta_1,(\beta_2,\beta_1)=0$$

则得

$$b_{12}=\frac{(\alpha_2,\beta_1)}{(\beta_1,\beta_1)}$$

同理，令

$$\beta_3=\alpha_3-b_{13}\beta_1-b_{23}\beta_2,(\beta_3,\beta_1)=0,(\beta_3,\beta_2)=0$$

则得

$$b_{13}=\frac{(\alpha_3,\beta_1)}{(\beta_1,\beta_1)},b_{23}=\frac{(\alpha_3,\beta_2)}{(\beta_2,\beta_2)}$$

依此类推，若 β_1、β_2、$\cdots$、β_k 已确定，$\beta_i\neq\theta,i\in\underline{k},k<m$ 且

$$(\beta_i,\beta_j)=0 \qquad i\neq j,i,j\in\underline{k}$$

令

$$\beta_{k+1}=\alpha_{k+1}-\sum_{i=1}^{k}b_{i,k+1}\beta_i$$

$$(\beta_{k+1},\beta_j)=0 \qquad i\in\underline{k}$$

则可得

$$b_{i,k+1}=\frac{(\alpha_{k+1},\beta_i)}{(\beta_i,\beta_i)} \qquad i\in\underline{k}$$

从而得正交向量组 $\beta_1,\beta_2,\cdots,\beta_{k+1}$，可一直进行到 $k+1=m$ 为

止。而

$$\alpha_i = \beta_i + \sum_{j=1}^{i-1} b_{ji}\beta_j \qquad i \in \underline{m}$$

即

$$[\alpha_1, \alpha_2, \cdots, \alpha_m] = [\beta_1, \beta_2, \cdots, \beta_m]B$$

其中 $B=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ & b_{22} & \cdots & b_{2m} \\ & & \ddots & \vdots \\ & & & b_{mm} \end{bmatrix}$

由于 $b_{ii}\neq 0, i\in \underline{m}$，故 B 是可逆上三角阵。

若再令

$$\gamma_i = \frac{\beta_i}{\|\beta_i\|} \qquad i \in \underline{m}$$

则 $[\gamma_1, \gamma_2, \cdots, \gamma_m]$ 即为 $V_n(C,U)(V_n(R,E))$ 的标准正交向量组。

定理 3 所介绍的求正交向量组的方法，成为**向量组的施密特(Schmidt)正交化方法**。

例 4 在 R^4 中，取

$$(x, y) = x^T y \qquad \forall\ x, y \in R^4$$

试将 R^4 中的线性无关向量组

$\alpha_1=(1,1,0,0)^T, \alpha_2=(1,0,1,0)^T,$

$\alpha_3=(-1,0,0,1)^T, \alpha_4=(1,-1,-1,1)^T$

化为 R^4 的标准正交基。

解 先正交化。为此取 $\beta_1=\alpha_1$，并计算

$$b_{12} = \frac{(\alpha_2, \beta_1)}{(\beta_1, \beta_1)} = \frac{1}{2}$$

于是

$$\beta_2 = \alpha_2 - b_{12}\beta_1 = \left(\frac{1}{2}, -\frac{1}{2}, 1, 0\right)^T$$

再由

$$b_{i3} = \frac{(\alpha_3, \beta_i)}{(\beta_i, \beta_i)} \qquad i \in \underline{2}$$

得 $b_{13} = -\frac{1}{2}$，$b_{23} = -\frac{1}{3}$，于是

$$\beta_3 = \alpha_3 - b_{13}\beta_1 - b_{23}\beta_2 = \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1\right)^T$$

最后由

$$b_{i4} = \frac{(\alpha_4, \beta_i)}{(\beta_i, \beta_i)} \qquad i \in \underline{3}$$

得 $b_{14} = b_{24} = b_{34} = 0$，即

$$\beta_4 = \alpha_4 = (1, -1, -1, 1)^T$$

再标准化，得

$$\gamma_1 = \frac{\beta_1}{\|\beta_1\|} = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0, 0\right)^T$$

$$\gamma_2 = \frac{\beta_2}{\|\beta_2\|} = \left(\frac{\sqrt{6}}{6}, -\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{3}, 0\right)^T$$

$$\gamma_3 = \frac{\beta_3}{\|\beta_3\|} = \left(-\frac{\sqrt{3}}{6}, \frac{\sqrt{3}}{6}, \frac{\sqrt{3}}{6}, \frac{\sqrt{3}}{2}\right)^T$$

$$\gamma_4 = \frac{\beta_4}{\|\beta_4\|} = \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)^T$$

于是$[\gamma_1, \gamma_2, \gamma_3, \gamma_4]$为 R^4 的一个标准正交基。

§2.3 正交子空间

这里我们将向量间的正交概念再推广到子空间之间，并讨论如何将酉(欧氏)空间分解为彼此正交的子空间的和。

定义 1 设 V_1 是 $V_n(C, U)$($V_n(R, E)$)的子空间，$x \in$

$V_n(C,U)$（$V_n(R,E)$）是某一向量，若

$$(x,y)=0 \qquad \forall\, y\in V_1$$

则称**向量 x 与子空间 V_1 正交**，记为 $x\perp V_1$。

定义 2 设 V_1、V_2 是 $V_n(C,U)$（$V_n(R,E)$）的子空间，若

$$x\perp V_2 \qquad \forall\, x\in V_1$$

则称**子空间 V_1、V_2 正交**，记为 $V_1\perp V_2$。

例 1 设 $e_i, i\in \underline{4}$ 是 $V_4(R,E)$ 的标准正交基，$V_1=\mathrm{span}(e_1,e_2)$，则 $e_3\perp V_1$。又若 $V_2=\mathrm{span}(e_3,e_4)$，则 $V_1\perp V_2$。

酉(欧氏)空间的正交子空间之间，有如下性质。

定理 1 设 V_1、V_2 是 $V_n(C,U)$（$V_n(R,E)$）的子空间，且 $V_1\perp V_2$，则

a） $V_1\cap V_2=\{\theta\}$

b） $\dim(V_1+V_2)=\dim V_1+\dim V_2$

定理成立是显然的。

定理 2 设 $A\in C^{n\times m}(R^{n\times m})$，$B\in C^{n\times s}(R^{n\times s})$，则 $R(A)\perp R(B)$ 的充要条件为

$$A^HB=O(A^TB=O)$$

证 设 $A=(\alpha_1,\alpha_2,\cdots,\alpha_m)$，$B=(\beta_1,\beta_2,\cdots,\beta_s)$，其中 α_i，$\beta_j\in C^n(R^n)$，$i\in \underline{m}$，$j\in \underline{s}$，于是 $A^HB=O(A^TB=O)$ 等价于

$$(\alpha_i,\beta_j)=0 \qquad i\in \underline{m}, j\in \underline{s} \tag{1}$$

任取

$x=(\alpha_1,\alpha_2,\cdots,\alpha_m)a\in R(A)$，$a=(a_1,a_2,\cdots,a_m)^T\in C^m(R^m)$

$y=(\beta_1,\beta_2,\cdots,\beta_s)b\in R(B)$，$b=(b_1,b_2,\cdots,b_s)^T\in C^s(R^s)$

则(1)式等价于

$$(x,y)=\sum_{i=1}^{m}\sum_{j=1}^{s}\overline{a_i}b_j(\alpha_i,\beta_j)=0$$

即 $R(A)\perp R(B)$。

推论 设 $A\in C^{m\times n}(R^{m\times n})$，则

a) $R(A)\perp N(A^H)$

b) $N(A)\perp R(A^H)$

证 若 $x\in N(A^H)$，则 $A^Hx=\theta$；$y\in R(A)$，则存在 z，使 $y=Az$，因

$$(x,y)=x^Hy=x^HAz=(A^Hx)^Hz=0$$

于是 $R(A)\perp N(A^H)$，而 b)是 a)中用 A 代替 A^H 所得的结果。

定义 3 设 V_1、V_2 是 $V_n(C,U)(V_n(R,E))$ 的子空间，且 $V_1\perp V_2$，若

$$V_1+V_2=V_n(C,U)(V_n(R,E))$$

则称 V_1 与 V_2 **互为正交补空间**，记为 $V_1=(V_2)_\perp$ 或 $V_2=(V_1)_\perp$ 或

$$V_n(C,U)=V_1\oplus V_2\quad(V_n(R,E)=V_1\oplus V_2)$$

在上述推论中，显然

$$\dim R(A)+\dim N(A^H)=m$$

$$\dim R(A^H)+\dim N(A)=n$$

于是

$$C^m=R(A)\oplus N(A^H)$$

$$C^n=R(A^H)\oplus N(A)$$

在第一章，已知线性空间的直和分解不唯一，但酉(欧氏)空间的正交分解是唯一的，即给定 V_1 后，正交补 V_2 是唯一的。证明如下。

定理 3 设 V_1 是 $V_n(C,U)(V_n(R,E))$ 的任一子空间，则存在唯一的子空间 $V_2\subset V_n(C,U)(V_n(R,E))$，使

$$V_1\oplus V_2=V_n(C,U)(V_n(R,E))$$

证 设 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_r$ 为 V_1 的标准正交基，经扩充后，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_r,\varepsilon_{r+1},\cdots,\varepsilon_n$ 为 $V_n(C,U)(V_n(R,E))$ 的标准正交基，若

$$V_2 = \text{span}(\varepsilon_{r+1},\varepsilon_{r+2},\cdots,\varepsilon_n)$$

显然 $V_1 \perp V_2$，且

$$V_1 + V_2 = V_n(C,U)(V_n(R,E))$$

故

$$V_1 \oplus V_2 = V_n(C,U)(V_n(R,E))$$

再证唯一性。若另有 $V_n(C,U)(V_n(R,E))$ 的子空间 V_3，使

$$V_1 \oplus V_3 = V_n(C,U)(V_n(R,E))$$

则对任意 $\theta \neq \beta \in V_3$，有 $\beta \notin V_1$，且

$$(\alpha,\beta) = 0 \qquad \forall \alpha \in V_1$$

所以 $\beta \in V_2$，$V_3 \subset V_2$，同理可证 $V_2 \subset V_3$，故 $V_2 = V_3$。

定理 3 的证明，给出了求正交补空间的具体方法。下面再讨论正交补空间的性质。

定理 4 设 S、Q 为 $V_n(C,U)(V_n(R,E))$ 的子空间，则

a) $(S+Q)_\perp = S_\perp \cap Q_\perp$；

b) $(S\cap Q)_\perp = S_\perp + Q_\perp$

证 *a*) 设 $x \in (S+Q)_\perp$，当且仅当 $x \in S_\perp$ 且 $x \in Q_\perp$，即 $x \in S_\perp \cap Q_\perp$。

b) 由 *a*) 有

$$(S\cap Q)_\perp = [(S_\perp + Q_\perp)_\perp]_\perp = S_\perp + Q_\perp$$

§ 2.4 酉(欧氏)空间的几种映射

这里将介绍酉(欧氏)空间中几种最常用的映射：酉(正交)变换、次酉(次正交)映射和正交投影。

定义 1 若 $V_n(C,U)(V_n(R,E))$ 的变换 σ 满足

$$(\sigma(x),\sigma(y)) = (x,y)$$
$$\forall\, x,y \in V_n(C,U)(V_n(R,E)).$$

则称 σ 为 $V_n(C,U)(V_n(R,E))$的**酉(正交)变换**。

例 1 设 $H=I_n-2uu^H\in C^{n\times n}$，$u\in C^n$，且有 $u^Hu=1$，则 H 是 $C^n(U)$中的酉变换。

解 对任意的 $x,y\in C^n(U)$，有

$$(Hx,Hy) = (x-2uu^Hx, y-2uu^Hy) = (x^H-2x^Huu^H)$$
$$\times(y-2uu^Hy) = x^Hy = (x,y)$$

故 H 是 $C^n(U)$中的酉变换。

H 称为**浩斯何尔德(Householder)镜象变换矩阵**，它将 x 映射为关于与 u 正交的 $n-1$ 维子空间的镜象。

例 2 设

$$A=\begin{bmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & \cos\theta & & & & -\sin\theta & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & -\sin\theta & & & & \cos\theta & & \\ & & & & & & & & 1 & \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1 \end{bmatrix}_{n\times n}$$

称为吉文斯(Givens)矩阵，试证 A 是 $R^n(E)$的正交变换。

证 易知 $A^TA=I_n$，且

$$(Ax,Ay) = x^TA^TAy = x^Ty = (x,y) \quad \forall\, x,y \in R^n(E)$$

故 A 是 $R^n(E)$的正交变换。

定理 1 $V_n(C,U)(V_n(R,E))$的酉(正交)变换是线性变换。

证 由假设知

$$(\sigma(x),\sigma(y)) = (x,y)$$
$$\forall\ x,y \in V_n(C,U)(V_n(R,E))$$

于是有

$$(\sigma(x+y)-\sigma(x)-\sigma(y),\sigma(x+y)-\sigma(x)-\sigma(y)) = 0$$

即

$$\sigma(x+y) = \sigma(x)+\sigma(y)$$

又

$$(\sigma(\lambda x)-\lambda\sigma(x),\sigma(\lambda x)-\lambda\sigma(x)) = 0 \qquad \forall\ \lambda \in C(R)$$

即

$$\sigma(\lambda x) = \lambda\sigma(x)$$

故 σ 是 $V_n(C,U)(V_n(R,E))$的线性变换。

定理 2 设 σ 是 $V_n(C,U)(V_n(R,E))$的线性变换,则下列条件等价:

a) σ 为酉(正交)变换;

b) $\|\sigma(x)\| = \|x\|$, $\forall\ x\in V_n(C,U)(V_n(R,E))$

c) 当 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 为 $V_n(C,U)(V_n(R,E))$的标准正交基时,$\sigma(\varepsilon_1),\sigma(\varepsilon_2),\cdots,\sigma(\varepsilon_n)$也是它的标准正交基;

d) σ 在任一标准正交基下的矩阵是酉(正交)矩阵。

证 $a)\Rightarrow b)$显然;

$b)\Rightarrow a)$,由 $b)$有

$$(\sigma(x+y),\sigma(x+y))=(x+y,x+y)$$
$$(\sigma(x+iy),\sigma(x+iy))=(x+iy,x+iy)$$
$$(i=\sqrt{-1})$$

上二式展开得

$$(\sigma(x),\sigma(y))+(\sigma(y),\sigma(x))=(x,y)+(y,x)$$

$$(\sigma(x),\sigma(y))-(\sigma(y),\sigma(x))=(x,y)-(y,x)$$

相加得 a)；

$a)\Rightarrow c)$　由 $(\varepsilon_i,\varepsilon_j)=\delta_{ij}$　　$i,j\in \underline{n}$

及

$$(\sigma(x),\sigma(y)) = (x,y)$$

有

$$(\sigma(\varepsilon_i),\sigma(\varepsilon_j)) = (\varepsilon_i,\varepsilon_j) = \delta_{ij} \qquad i,j \in \underline{n}$$

即 c)成立；

$c)\Rightarrow a)$　设

$$(\varepsilon_i,\varepsilon_j) = (\sigma(\varepsilon_i),\sigma(\varepsilon_j)) = \delta_{ij} \qquad i,j \in \underline{n}$$

任取

$$x=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\tilde{x}\in V_n(C,U)(V_n(R,E))$$

$$y=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\tilde{y}\in V_n(C,U)(V_n(R,E))$$

其中　$\tilde{x}=(x_1,x_2,\cdots,x_n)^T,\tilde{y}=(y_1,y_2,\cdots,y_n)^T\in C^n(R^n)$

有

$$\begin{aligned}(\sigma(x),\sigma(y))&=\left(\sum_{i=1}^{n}x_i\sigma(\varepsilon_i),\sum_{j=1}^{n}y_j\sigma(\varepsilon_j)\right)\\&=\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{x_i}y_j(\sigma(\varepsilon_i),\sigma(\varepsilon_j))\\&=\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{x_i}y_j(\varepsilon_i,\varepsilon_j)=(x,y)\end{aligned}$$

即 a)成立；

$c)\Rightarrow d)$设

$$\sigma(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)A$$

且 $A=(\alpha_1,\alpha_2,\cdots,\alpha_n)$，则$(\sigma(\varepsilon_i),\sigma(\varepsilon_j))=\delta_{ij}(i,j\in \underline{n}\Longleftrightarrow\alpha_i^H\alpha_j=\delta_{ij}(\alpha_i^T\alpha_j=\delta_{ij})(i,j\in\underline{n}\Longleftrightarrow A$ 是酉(正交)矩阵。

此定理说明，$V_n(C,U)(V_n(R,E))$的酉(正交)变换与

$C^{n\times n}(R^{n\times n})$的酉(正交)矩阵相对应，故酉(正交)矩阵亦将具有相应的性质。

定理 3 设 $A\in C^{n\times n}(R^{n\times n})$是酉(正交)矩阵，则

a) $(Ax,Ay)=(x,y) \quad \forall\ x,y\in C^n(U)(R^n(E))$

b) $\|Ax\|=\|x\| \quad \forall\ x\in C^n(U)(R^n(E))$

c) $A^H(A^T)$也是酉(正交)矩阵；

d) 若 $B\in C^{n\times n}(R^{n\times n})$也是酉(正交)矩阵，则 AB、BA 也是酉(正交)矩阵；

e) 酉(正交)矩阵的特征值的模(绝对值)为 1。

证 *a*) $(Ax,Ay)=x^HA^HAy=x^Hy=(x,y)$

b) 由 *a*)即得；

c) 因$(A^H)^HA^H=(AA^H)^H=(I_n)^H=I_n$

d) $(AB)^H(AB)=B^HA^HAB=B^HB=I_n$

$(BA)^H(BA)=A^HB^HBA=A^HA=I_n$

e) 设 $Ax=\lambda x$，则

$$\bar{\lambda}x^Hx=(Ax,x)=x^HA^Hx=x^HA^{-1}x$$
$$=\frac{1}{\lambda}x^Hx$$

故$\bar{\lambda}\lambda=1$，$|\lambda|=1$。

酉(正交)变换保持长度不变，如果只要求在部分空间上具有这种特性，则可引出条件相对弱一些的次酉(次正交)映射。

定义 2 设 $\sigma:V_m(C,U_1)\to V_n(C,U_2)(V_m(R,E_1)\to(V_n(R,E_2))$，且对任意的 $x,y\in S\subset V_m(C,U_1)(V_m(R,E_1))$，$S=(\mathrm{Ker}\sigma)_\perp$，有

$$(\sigma(x),\sigma(y))_2=(x,y)_1$$

则称 σ 为由 $V_m(C,U_1)$到 $V_n(C,U_2)$(由 $V_m(R,E_1)$到 $V_n(R,$

E_2))的次酉(次正交)映射。

例3 设

$$A=\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{1}{\sqrt{2}} & 0\\ 0 & \frac{1}{\sqrt{2}} & 0\end{bmatrix}$$

则 A 是 $C^3(U)$ 的次酉映射。

解 因为

$$A^H=\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ 0 & 0 & 0\end{bmatrix}$$

显然 $R(A^H)=\mathrm{span}(e_1,e_2)$，其中 $e_1=(1,0,0)^T$，$e_2=(0,1,0)^T$，则对任意的 $x=(x_1,x_2,0)^T$，$y=(y_1,y_2,0)^T\in R(A^H)$，有

$$(Ax,Ay)=x^HA^HAy=x^H\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}y$$

$$=x^Hy=(x,y)$$

故 A 是 $C^3(U)$ 的次酉映射(次酉变换)。

次酉(次正交)映射有与酉(正交)变换相类似的性质。

最后，再介绍另一类重要映射——正交投影。由于酉(欧氏)空间分解为两个互为正交补空间的子空间的和是唯一的，设

$$S\oplus T=V_n(C,U)(V_n(R,E))$$

今取 $x\in V_n(C,U)(V_n(R,E))$，且

$$x=x_1+x_2\qquad x_1\in S,x_2\in T$$

则 $x_i, i\in\underline{2}$ 与 x 也是一一对应的，从而引出正交投影的概念。

定义 3 设 $S\oplus T=V_n(C,U)(V_n(R,E))$，$\sigma: V_n(C,U)(V_n(R,E))\to S$ 是一个映射，若 $x=x_1+x_2$，$x_1\in S$，$x_2\in T$，且 $\sigma(x)=x_1$，则称 σ 为由 $V_n(C,U)(V_n(R,E))$ 到 S 的**正交投影**，记为 P_s。

正交投影有下面的性质。

定理 4 设 P_s 为 $C^n(U)(R^n(E))$ 的正交投影，$S=R(U_1)$，$U_1=(u_1,u_2,\cdots,u_r)$，其中 $u_1,u_2,\cdots,u_r$ 为 S 的标准正交基，则

$$P_s=U_1U_1^H$$

证 设 $u_1,u_2,\cdots,u_r,u_{r+1},\cdots,u_n$ 为 $C^n(U)(R^n(E))$ 的标准正交基，$U_2=(u_{r+1},u_{r+2},\cdots,u_n)$，$U=(U_1,U_2)$，则

$$UU^H=(U_1U_2)\begin{bmatrix}U_1^H\\U_2^H\end{bmatrix}=U_1U_1^H+U_2U_2^H=I_n$$

且

$$R(U_1)\oplus R(U_2)=C^n(U)(R^n(E))$$

任取 $x\in C^n(U)(R^n(E))$

$$x=U_1U_1^Hx+U_2U_2^Hx$$

显然 $U_1U_1^Hx\in R(U_1)$，$U_2U_2^Hx\in R(U_2)$，于是

$$P_sx=U_1U_1^Hx$$

由 x 的任意性，有

$$P_s=U_1U_1^H$$

此定理说明，在 $C^n(U)(R^n(E))$ 中，正交投影是一个矩阵，这个矩阵有下述性质。

定理 5 在 $C^n(U)(R^n(E))$ 中，矩阵 E 为正交投影的充要条件为

$$E = E^H = E^2 (E = E^T = E^2)$$

证 必要性：设 E 为 $C^n(U)(R^n(E))$ 到其子空间 S 的正交投影，由定理 4，若 U_1 为由 S 的标准正交基组成的矩阵，则

$$E = U_1 U_1^H$$

显然

$$E^H = (U_1 U_1^H)^H = U_1 U_1^H = E$$

$$E^2 = U_1 U_1^H U_1 U_1^H = U_1 I_r U_1^H = U_1 U_1^H = E$$

充分性：设 $E = E^H = E^2$，$R(U_1) = R(E)$，$U_1 = (u_1, u_2, \cdots, u_r)$，$U_1^H U_1 = I_r$，$\text{rank}E = r$，故存在 $V_1^H \in C^{r\times n}(R^{r\times n})$，使

$$E = U_1 V_1^H \qquad \text{rank}V_1 = r$$

于是

$$U_1 V_1^H = V_1 U_1^H = V_1 U_1^H U_1 V_1^H = V_1 V_1^H$$

得

$$(U_1 - V_1) V_1^H = O$$

考虑到 V_1^H 具有线性无关行，则 $U_1 = V_1$，于是 E 为正交投影。

习 题 二

一、设 $\{e_1, e_2, \cdots, e_n\}$ 为 $V_n(R)$ 的一组基，$x, y \in V_n(R)$，且 $x = \sum_{i=1}^{n} \alpha_i e_i$，$y = \sum_{i=1}^{n} \beta_i e_i$，试证

$$(x, y) = \sum_{i=1}^{n} i \alpha_i \beta_i$$

为 $V_n(R)$ 的内积。

二、若 $A = -A^T \in R^{n\times n}$，则

$$x^T A x = 0 \qquad \forall x \in R^n$$

三、设 $A \in C^{m\times n}$，则

1. $N(A) = N(A^H A)$，$N(A^H) = N(A A^H)$

2. $R(A)=R(AA^H)$, $R(A^H)=R(A^HA)$

3. $\text{rank}A=\text{rank}A^HA=\text{rank}AA^H$

4. 方程

$$A^HAx = A^Hb \qquad \forall A \in C^{m\times n}, \forall b \in C^m$$

一定有解。

四、由标准正交基到标准正交基的过渡矩阵一定是酉(正交)矩阵。

五、若 $A\in C^{n\times n}$ 为次酉阵，则 A 的非零特征值的模为 1；A^HA 的特征值非零即 1。

六、设 $V_n(C,U)=S\oplus T$，P_s 为由 $V_n(C,U)$ 到 S 的正交投影，则 $V_n(C,U)$ 到 T 的正交投影为 I_n-P_s。

七、设 $A\in C^{n\times n}$ 是正交投影，则 A 的特征值非 0 即 1。

八、若 $A=A^H=A^2\in C^{n\times n}$，$\text{rank}A=r$，则存在 $V\in U^{n\times n}$，使

$$V^HAV = \text{diag}(I_r,O)$$

第三章　矩阵的分解

把矩阵分解为具有某种特性的矩阵之积或矩阵的和，在矩阵理论的研究与应用中，都是十分重要的。本章仅就一些最常用的矩阵分解作简要介绍。

§3.1　n 阶方阵的三角分解

定理 1　若 $A\in C^{n\times n}$ 的顺序主子式全不为零，则 A 可唯一地分解成

$$A=L\tilde{U}=(\tilde{L}U) \tag{1}$$

其中　L 为下三角阵；$\tilde{U}$ 为单位上三角阵（$\tilde{L}$ 为单位下三角阵，U 为上三角阵）。

矩阵的这种分解称为**三角分解或克劳特（Crout）分解**。它也可表为

$$A=\tilde{L}D\tilde{U}$$

其中　D 为对角阵。

定理的证明略去。式(1)的具体做法如下：

设 $L=(l_{ij})$，$\tilde{U}=(u_{ij})$，i，$j\in\underline{n}$，比较(1)式两端的对应元素，得

$$l_{i1}=\alpha_{i1} \qquad i\in\underline{n} \tag{2}$$

$$u_{1j}=\frac{\alpha_{1j}}{\alpha_{11}} \qquad j=2,3,\cdots,n \tag{3}$$

$$l_{ik}=a_{ik}-\sum_{r=1}^{k-1} l_{ir}u_{rk} \quad k\leqslant i,\ i\in \underline{n} \tag{4}$$

$$u_{kj}=\frac{a_{kj}-\sum\limits_{r=1}^{k-1} l_{kr}u_{rj}}{l_{kk}} \quad \begin{array}{l} k=2,3,\cdots,n-1 \\ j=k+1,k+2,\cdots,n \end{array} \tag{5}$$

n 阶方阵 A 的三角分解，在求解线性非齐次方程组时十分方便。设有方程组

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \tag{6}$$

即

$$L\tilde{U}\boldsymbol{x}=\boldsymbol{b}$$

令 $\tilde{U}\boldsymbol{x}=\boldsymbol{y}$，则(6)相当于

$$\begin{cases} L\boldsymbol{y}=\boldsymbol{b} \\ \tilde{U}\boldsymbol{x}=\boldsymbol{y} \end{cases} \tag{7}$$

即先解(7)中第一式得 $\boldsymbol{y}$，然后代入(7)的第二式，得 $\boldsymbol{x}$ 即为所求的解。由于(7)的两个方程的系数矩阵均为三角阵，所以求解很方便。

例 用三角分解求解方程组

$$\begin{cases} 2x_1+x_2-5x_3+x_4=8 \\ x_1-3x_2-6x_4=9 \\ 2x_2-x_3+2x_4=5 \\ x_1+4x_2-7x_3+6x_4=0 \end{cases}$$

解 记 $\boldsymbol{x}=(x_1,x_2,x_3,x_4)^T$，$\boldsymbol{b}=(8,9,5,0)^T$。

$$\boldsymbol{A}=\begin{bmatrix} 2 & 1 & -5 & 1 \\ 1 & -3 & 0 & -6 \\ 0 & 2 & -1 & 2 \\ 1 & 4 & -7 & 6 \end{bmatrix}$$

则由式(2)得

$$l_{11}=2,\ l_{21}=1,\ l_{31}=0,\ l_{41}=1$$

由公式(3)得

$$u_{12}=\frac{1}{2},\quad u_{13}=-\frac{5}{2},\quad u_{14}=\frac{1}{2}$$

由公式(4)得

$$l_{22}=-\frac{7}{2},\quad l_{32}=2,\quad l_{42}=\frac{7}{2}$$

由公式(5)得

$$u_{23}=-\frac{5}{7},\quad u_{24}=\frac{13}{7}$$

再由公式(4)得

$$l_{33}=\frac{3}{7},\quad l_{43}=-2$$

再由公式(5)得

$$u_{34}=-4$$

最后由公式(4)得

$$l_{44}=-9$$

于是得

$$A=\begin{bmatrix}2 & & & \\ 1 & -\frac{7}{2} & & \\ 0 & 2 & \frac{3}{7} & \\ 1 & \frac{7}{2} & -2 & -9\end{bmatrix}\begin{bmatrix}2 & \frac{1}{2} & -\frac{5}{2} & \frac{1}{2} \\ & 1 & -\frac{5}{7} & \frac{13}{7} \\ & & 1 & -4 \\ & & & 1\end{bmatrix}$$

代入(7)式的第一式，求得 $y=\left(4,\ -\frac{10}{7},\ -5,\ 1\right)^T$，再将此

结果代入(7)式的第二式得

$$x=(3, -4, -1, 1)^T$$

此即所求方程组的解。

§ 3.2 n 阶方阵的约当(Jordan)标准形

已知相似矩阵有相同的特征值，特征值的计算是比较复杂的，如果在相似矩阵中，能找到结构简单的矩阵，显然，这将会给矩阵特征值理论的研究带来方便。

首先，证明任一 n 阶方阵均能与上三角阵酉相似。

定理 1 设 $A\in C^{n\times n}$，则存在 $U\in U^{n\times n}$(这里 $U^{n\times n}$表示 n 阶酉阵的集合)，使得

$$A=URU^H \tag{1}$$

其中 R 为对角线元是 A 的特征值的上三角矩阵。

这个定理称为司楚尔(Schur)定理。

证 用数学归纳法，显然 $n=1$ 时，定理成立。设 $n=m$ 时定理成立，证明 $n=m+1$ 时定理成立。

取 $\theta\neq u_1\in C^{m+1}$，使

$$Au_1=\lambda u_1 \qquad \|u_1\|=1$$

再取 $u_2, u_3, \cdots, u_{m+1}\in C^{m+1}$，使 $u_1, u_2, \cdots, u_{m+1}$为 C^{m+1}的标准正交基，记 $U_1=(u_1, u_2, \cdots, u_{m+1})\in U^{(m+1)\times(m+1)}$于是

$$U_1^H AU_1=\begin{pmatrix} \lambda & c^T \\ O & A_1 \end{pmatrix}$$

其中 $c\in C^m$，$A_1\in C^{m\times m}$。

由归纳法假设知，存在 $V_1\in U^{m\times m}$，使

$$V_1^H A_1V_1=R_1$$

其中 R_1为对角线元是 A_1的特征值的上三角阵。

显然，A_1的特征值也是 A 的特征值，再令

$$U_2=\begin{pmatrix}1 & O\\ O & V_1\end{pmatrix},\quad U=U_1U_2$$

则

$$\begin{aligned}U^HAU&=U_2{}^HU_1{}^HAU_1U_2\\&=\begin{pmatrix}1 & O\\ O & V_1{}^H\end{pmatrix}\begin{pmatrix}\lambda & c^T\\ O & A_1\end{pmatrix}\begin{pmatrix}1 & O\\ O & V_1\end{pmatrix}\\&=\begin{pmatrix}\lambda & c^TV_1\\ O & V_1{}^HA_1V_1\end{pmatrix}=\begin{pmatrix}\lambda & c^TV_1\\ O & R_1\end{pmatrix}=R\end{aligned}$$

易知R为对角线元是A的特征值的上三角矩阵，即(1)式成立。

下面再进一步讨论上三角阵R的结构特征，得知一般情况下，R是一个对角块阵，而且那些对角块都是“二对角”阵，这就是将要介绍的矩阵的约当标准形，为了讲清约当标准形的结构特征，先介绍用来表示这些结构特征的一些基本概念。

定义1 矩阵$A\in C^{n\times n}$的特征矩阵$\lambda I-A$的k阶子式的最大公因式$D_k(\lambda)$称为$\lambda I-A$的**k阶行列式因子**。

显然$D_{k-1}(\lambda)\,|\,D_k(\lambda)$，它表示$\lambda$的多项式$D_{k-1}(\lambda)$整除$\lambda$的多项式$D_k(\lambda)$。

$d_k(\lambda)=\dfrac{D_k(\lambda)}{D_{k-1}(\lambda)}$称为$\lambda I-A$的**不变因子**(这里$d_1(\lambda)=D_1(\lambda)$或说$D_0(\lambda)=1$)。

为了确定$D_k(\lambda)$的唯一性，假定$D_k(\lambda)$是首1的，于是$D_n(\lambda)=\det(\lambda I-A)$，$D_k(\lambda)=d_1(\lambda)d_2(\lambda)\cdots d_k(\lambda)$，$k\in\underline{n}$，相似矩阵有相同的行列式因子，且$d_{k-1}(\lambda)\,|\,d_k(\lambda)$(证略)。

在复数域内，将不变因子分解为λ的一次因式的幂积。设λ_1、λ_2、…，λ_s为A的相异特征值，则

$$d_1(\lambda)=(\lambda-\lambda_1)^{k_{11}}(\lambda-\lambda_2)^{k_{12}}\cdots(\lambda-\lambda_\sigma)^{k_{1\sigma}}$$
$$d_2(\lambda)=(\lambda-\lambda_1)^{k_{21}}(\lambda-\lambda_2)^{k_{22}}\cdots(\lambda-\lambda_\sigma)^{k_{2\sigma}}$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$d_n(\lambda)=(\lambda-\lambda_1)^{k_{n1}}(\lambda-\lambda_2)^{k_{n2}}\cdots(\lambda-\lambda_\sigma)^{k_{n\sigma}}$$

其中 $\quad k_{1j}\leqslant k_{2j}\leqslant\cdots\leqslant k_{nj} \qquad j\in\underline{\sigma}$

这些一次因式称为 $\lambda I-A$ 的初等因子。显然，相似矩阵的初等因子相同。

例 1 求矩阵

$$A=\begin{bmatrix} \lambda_0 & 1 & & & \\ & \lambda_0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_0 \end{bmatrix}_{m\times m}$$

的特征矩阵 $\lambda I-A$ 的初等因子。

解 A 的特征矩阵

$$\lambda I-A=\begin{bmatrix} \lambda-\lambda_0 & -1 & & & \\ & \lambda-\lambda_0 & -1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & -1 \\ & & & & \lambda-\lambda_0 \end{bmatrix}$$

易知 $D_m(\lambda)=(\lambda-\lambda_0)^m$，由于 $\det(\lambda I-A)$ 中有一个 $m-1$ 阶子式为

$$\begin{vmatrix} -1 & 0 & \cdots\cdots & & 0 \\ \lambda-\lambda_0 & -1 & \ddots & & \vdots \\ 0 & \lambda-\lambda_0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0\cdots\cdots & 0 & & \lambda-\lambda_0 & -1 \end{vmatrix}=(-1)^{m-1}$$

故 $D_{m-1}(\lambda)=1$，于是

$$D_{m-2}(\lambda)=D_{m-3}(\lambda)=\cdots=D_2(\lambda)=D_1(\lambda)=1$$

由此可知，$\lambda I-A$ 的不变因子

$$d_1(\lambda)=d_2(\lambda)=\cdots=d_{m-1}(\lambda)=1$$

$$d_m(\lambda)=(\lambda-\lambda_0)^m$$

因此 $\lambda I-A$ 的初等因子为 $(\lambda-\lambda_0)^m$。

例 2 求矩阵

$$A=\begin{bmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{bmatrix}$$

的特征矩阵 $\lambda I-A$ 的初等因子。

解 这里采用对 $\lambda I-A$ 进行“初等变换”的办法，求出与 $\lambda I-A$“等价”的 λ 矩阵

$$\begin{bmatrix} d_1(\lambda) & & & \\ & d_2(\lambda) & & \\ & & \ddots & \\ & & & d_n(\lambda) \end{bmatrix}$$

其中 $d_1(\lambda)$、$d_2(\lambda)$、…、$d_n(\lambda)$ 就是我们要找的 $\lambda I-A$ 的不变因子，由此即可求出 $\lambda I-A$ 的初等因子。

$$\lambda I-A=\begin{bmatrix} \lambda-4 & -6 & 0 \\ 3 & \lambda+5 & 0 \\ 3 & 6 & \lambda-1 \end{bmatrix}\to\begin{bmatrix} 1 & 2 & \frac{\lambda-1}{3} \\ 3 & \lambda+5 & 0 \\ \lambda-4 & -6 & 0 \end{bmatrix}$$

$$\to\begin{bmatrix} 1 & 2 & \frac{\lambda-1}{3} \\ 0 & \lambda-1 & -(\lambda-1) \\ 0 & -2(\lambda-1) & -\frac{1}{3}(\lambda-1)(\lambda-4) \end{bmatrix}$$

$$
\rightarrow\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda-1 & -(\lambda-1) \\ 0 & -2(\lambda-1) & -\frac{1}{3}(\lambda-1)(\lambda-4) \end{bmatrix}
$$

$$
\rightarrow\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda-1 & -(\lambda-1) \\ 0 & 0 & -\frac{1}{3}(\lambda-1)(\lambda+2) \end{bmatrix}
$$

$$
\rightarrow\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda-1 & 0 \\ 0 & 0 & (\lambda-1)(\lambda+2) \end{bmatrix}
$$

故 $\lambda I-A$ 的初等因子为 $\lambda-1$、$\lambda-1$、$\lambda+2$。

定义 2 设 λ_i，$i\in\underline{\sigma}$ 为 $A\in C^{n\times n}$ 的相异特征值，且

$$
\det(\lambda I-A)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_\sigma)^{m_\sigma}
$$

$\sum_{i=1}^{\sigma} m_i=n$，则称 m_i 为 A 的特征值 λ_i 的**代数重复度**．若

$$
\operatorname{rank}(\lambda_i I-A)=n-\alpha_i \qquad i\in\underline{\sigma}
$$

则称 α_i，$i\in\underline{\sigma}$ 为 A 的特征值 λ_i 的**几何重复度**。

显然，特征值 λ_i 的代数重复度 m_i，即 λ_i 的重根的重数，而 λ_i 的几何重复度 α_i，即 λ_i 的特征向量子空间的维数。

在例 1 中，矩阵 A 的特征值 λ_0 的代数重复度为 m，而

$$
\operatorname{rank}(\lambda_0 I-A)=\operatorname{rank}\begin{bmatrix} 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & -1 \\ 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}=m-1
$$

故其几何重复度为 1。

在例 2 中，由于 $\det(\lambda I-A)=(\lambda-1)^2(\lambda+2)$，若记 $\lambda_1=1$，$\lambda_2=-2$，则 λ_1 的代数重复度为 2，λ_2 的代数重复度为 1。又

$$\operatorname{rank}(\lambda_1 I-A)=\operatorname{rank}\begin{bmatrix}-3 & -6 & 0\\ 3 & 6 & 0\\ 3 & 6 & 0\end{bmatrix}=1=3-2$$

$$\operatorname{rank}(\lambda_2 I-A)=\operatorname{rank}\begin{bmatrix}-6 & -6 & 0\\ 3 & 3 & 0\\ 3 & 6 & -3\end{bmatrix}=2=3-1$$

故 λ_1 的几何重复度为 2，λ_2 的几何重复度为 1。这里 λ_1、λ_2 的几何重复度与代数重复度相等。几何重复度与代数重复度之间的关系，有下述定理。

定理 2 设 λ_i，$i\in\underline{\sigma}$ 为 $A\in C^{n\times n}$ 的相异特征值，m_i、α_i 分别为 λ_i 的代数重复度与几何重复度，则

$$\alpha_i\leqslant m_i \qquad i\in\underline{\sigma}$$

证 由 m_i 的定义有

$$\det(\lambda I-A)=(\lambda-\lambda_i)^{m_i}g(\lambda) \qquad g(\lambda_i)\neq 0$$

又由 α_i 的定义可知

$$(\lambda_i I-A)x=\theta$$

有基础解系 ε_1，ε_2，$\cdots$，ε_{α_i}（即 λ_i 所对应的特征子空间的基）。由此可扩充使 ε_1，ε_2，$\cdots$，ε_{α_i}，ε_{α_i+1}，$\cdots$，ε_n 为 C^n 的基。由于 A 的特征子空间是 A 的不变子空间，故有

$$A(\varepsilon_0,\varepsilon_1,\cdots,\varepsilon_n)=(\varepsilon_0,\varepsilon_1,\cdots,\varepsilon_n)\left[\begin{array}{cccc|c}\lambda_i & & & & \\ & \lambda_i & & & A_1\\ & & \ddots & & \\ & & & \lambda_i & \\ \hline & & O & & A_2\end{array}\right]$$

其中　$A_1 \in C^{\alpha_i \times (n-\alpha_i)}$，$A_2 \in C^{(n-\alpha_i)\times(n-\alpha_i)}$。

记

$$T=(\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_n) \qquad B=\left[\begin{array}{cccc:c} \lambda_i & & & & \\ & \lambda_i & & & A_1 \\ & & \ddots & & \\ & & & \lambda_i & \\ \hdashline & & O & & A_2 \end{array}\right]$$

由于

$$\det(\lambda I_n - B) = (\lambda - \lambda_i)^{\alpha_i} \det(\lambda I_{n-\alpha_i} - A_2) = \det(\lambda I - A)$$

可知 $\alpha_i \leqslant m_i$，$i \in \underline{\sigma}$。

若矩阵 A 的每个特征值的代数重复度与几何重复度均相等，则称 A 为**单纯矩阵**，例 2 中的矩阵 A 就是单纯矩阵。

定理 3　单纯矩阵 A 与对角矩阵 Λ 相似，即

$$A = T\Lambda T^{-1}$$

其中　$\Lambda = \operatorname{diag}(\underbrace{\lambda_1, \lambda_1, \cdots, \lambda_1}_{m_1\text{个}}, \underbrace{\lambda_2, \cdots, \lambda_2}_{m_2\text{个}}, \cdots, \underbrace{\lambda_\sigma, \cdots, \lambda_\sigma}_{m_\sigma\text{个}})$ T 的列向量为相应于这些特征值的线性无关的特征向量。

只要注意到“属于每个特征值的线性无关的特征向量合起来也是线性无关的”这一事实，即知定理成立。

如果 A 不是单纯矩阵，那么 A 能与什么样的矩阵相似？下面定理 4 给了明确回答。

定理 4　设 $A \in C^{n\times n}$，则存在可逆矩阵 T，使得

$$A = TJT^{-1} \tag{2}$$

其中

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_\sigma \end{bmatrix}_{n\times n} \tag{3}$$

$$J_i=\begin{bmatrix} J_{i1} & & & \\ & J_{i2} & & \\ & & \ddots & \\ & & & J_{i\alpha_i} \end{bmatrix}_{m_i\times m_i} \qquad i\in \underline{\sigma} \tag{4}$$

$$J_{ik}=\begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}_{n_{ik}\times n_{ik}} \qquad i\in \underline{\sigma},\quad k\in \underline{\alpha_i} \tag{5}$$

其中$\lambda_i(i\in\underline{\sigma})$为$A$的相异特征值，$m_i$、$\alpha_i$分别为特征值$\lambda_i$的代数重复度与几何重复度，$n_{ik}$为初等因子$(\lambda-\lambda_i)^{n_{ik}}$的指数，$\sum_{k=1}^{\alpha_i} n_{ik}=m_i$。

证略。

J称为A的约当标准形，J_{ik}称为A的约当块。显然，定理3是定理4的特例。

例3 求矩阵

$$A=\begin{bmatrix} -2 & 2 & -1 \\ 0 & -2 & 0 \\ 1 & -4 & 0 \end{bmatrix}$$

的约当标准形。

解 由于

$$\det(\lambda I-A)=\begin{vmatrix} \lambda+2 & -2 & 1 \\ 0 & \lambda+2 & 0 \\ -1 & 4 & \lambda \end{vmatrix}=(\lambda+1)^2(\lambda+2)$$

可见$\lambda_1=-1$是二重特征值，$\lambda_2=-2$是特征方程的单根。又

$$\text{rank}(-1I-A)=\text{rank}\begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & 0 \\ -1 & 4 & -1 \end{bmatrix}=2=3-1$$

故 A 的约当标准形为

$$J=\begin{bmatrix}-1 & 1 & \\ & -1 & \\ & & -2\end{bmatrix}$$

例 4 求矩阵

$$A=\begin{bmatrix}3 & 1 & 0 & 0\\ -4 & -1 & 0 & 0\\ 0 & 0 & 2 & 1\\ 0 & 0 & -1 & 0\end{bmatrix}$$

的约当标准形。

解 A 是一对角块矩阵，记为

$$A=\begin{bmatrix}A_1 & \\ & A_2\end{bmatrix}$$

若记 A 的约当标准形的相应结构为

$$J=\begin{bmatrix}J_1 & \\ & J_2\end{bmatrix}$$

则 A_1 与 J_1 相似，A_2 与 J_2 相似，这样就可分别对低阶矩阵 A_1、A_2 求相应的约当标准形。

$$\lambda I_2-A_1=\begin{bmatrix}\lambda-3 & -1\\ 4 & \lambda+1\end{bmatrix}\to\begin{bmatrix}1 & \frac{\lambda+1}{4}\\ \lambda-3 & -1\end{bmatrix}$$

$$\to\begin{bmatrix}1 & \frac{\lambda+1}{4}\\ 0 & -\frac{1}{4}(\lambda-1)^2\end{bmatrix}\to\begin{bmatrix}1 & 0\\ 0 & (\lambda-1)^2\end{bmatrix}$$

$$\lambda I_2-A_2=\begin{bmatrix}\lambda-2 & -1\\ 1 & \lambda\end{bmatrix}\to\begin{bmatrix}1 & \lambda\\ \lambda-2 & -1\end{bmatrix}$$

$$\rightarrow\begin{bmatrix}1 & \lambda \\ 0 & -(\lambda-1)^2\end{bmatrix}\rightarrow\begin{bmatrix}1 & 0 \\ 0 & (\lambda-1)^2\end{bmatrix}$$

故 A 的初等因子为$(\lambda-1)^2$、$(\lambda-1)^2$，于是由定理 4 知 A 的约当标准形为

$$J=\begin{bmatrix}1 & 1 & & \\ & 1 & & \\ & & 1 & 1 \\ & & & 1\end{bmatrix}$$

推论 设 $A=TJT^{-1}$，则

a) $A^p=TJ^pT^{-1}$，$f(A)=Tf(J)T^{-1}$

其中 $f(\lambda)$是 λ 的多项式。

b) $(\lambda I-A)^p=T(\lambda I-J)^pT^{-1}$

c) $\mathrm{rank}(\lambda I-A)^{p_0}=\mathrm{rank}(\lambda I-J)^{p_0}$

d) 若 $\mathrm{rank}(\lambda_i I-A)^{p_0}=\mathrm{rank}(\lambda_i I-A)^{p_0+1}$，且

$\mathrm{rank}(\lambda_i I-A)^p\neq\mathrm{rank}(\lambda_i I-A)^{p+1}\quad 1\leqslant p\leqslant p_0-1$

则

$$p_0=\max_k n_{ik}$$

证略。

在某些场合，不仅需要求出 A 的约当标准形 J，而且需要求出相似变换矩阵 T。显然，T 中只有 $\alpha_1+\alpha_2+\cdots+\alpha_s$ 个 A 的特征向量，其余 $n-(\alpha_1+\alpha_2+\cdots+\alpha_s)$个列向量如何求？

定义 3 设 $A\in C^{n\times n}$，λ_0是 A 的特征值，若

$$\begin{cases}(A-\lambda_0 I)^{k-1}x\neq\theta \\ (A-\lambda_0 I)^k x=\theta\end{cases}\qquad \theta\neq x\in C^n \tag{6}$$

则称 x 为 A 的属于 λ_0的秩为 k 的广义特征向量。

下面讨论相似变换T的构成及求法，为此，将T按J分块成

$$T=(T_1T_2\cdots T_\sigma)_{n\times n} \tag{7}$$

$$T_i=(T_{i1}\,T_{i2}\cdots T_{i\alpha_i})_{n\times m_i}\quad i\in\underline{\sigma} \tag{8}$$

$$T_{ik}=(t^1_{ik}t^2_{ik}\cdots t^{n_{ik}}_{ik})_{n\times n_{ik}}\quad i\in\underline{\sigma},\ k\in\underline{\alpha_i} \tag{9}$$

由此可见，每个T_{ik}中，第一个向量t^1_{ik}是A的属于λ_i的特征向量，其余t^2_{ik}、…、$t^{n_{ik}}_{ik}$均为A的属于λ_i的广义特征向量。

由$A=TJT^{-1}$有$AT=TJ$，从而有

$$AT_i=T_iJ_i$$

$$AT_{ik}=T_{ik}J_{ik}$$

于是

$$A\,t^1_{ik}=\lambda_i t^1_{ik} \tag{10}$$

$$\begin{cases} A\,t^2_{ik}=t^1_{ik}+\lambda_i t^2_{ik} \\ A\,t^3_{ik}=t^2_{ik}+\lambda_i t^3_{ik} \\ \cdots\ \cdots\ \cdots\ \cdots \\ A\,t^{n_{ik}}_{ik}=t^{n_{ik}-1}_{ik}+\lambda_i t^{n_{ik}}_{ik} \end{cases} \tag{11}$$

当t^1_{ik}由(10)式求出之后，代入(11)式可求出t^2_{ik}、…、$t^{n_{ik}}_{ik}$这些广义特征向量。

例 5 求例 3 中矩阵$A=TJT^{-1}$的相似变换矩阵T。

解 记$T=(t_1t_2t_3)$，于是，记$t_1=(t_{11},t_{21},t_{31})^T$，则

$$(-1I-A)t_1=\begin{bmatrix}1 & -2 & 1\\ 0 & 1 & 0\\ -1 & 4 & -1\end{bmatrix}\begin{bmatrix}t_{11}\\ t_{21}\\ t_{31}\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

得$t_{21}=0$ $t_{11}=-t_{31}$，故可令$t_1=(-1,0,1)^T$。又

$$(A-(-1)I)t_2=t_1$$

即当$t_2=(t_{12},t_{22},t_{32})^T$时有

$$\begin{bmatrix}-1 & 2 & -1\\ 0 & -1 & 0\\ 1 & -4 & 1\end{bmatrix}\begin{bmatrix}t_{12}\\ t_{22}\\ t_{32}\end{bmatrix}=\begin{bmatrix}-1\\0\\1\end{bmatrix}$$

解得 $t_{22}=0$，$t_{12}+t_{32}=1$，故可令 $t_2=(0,0,1)^T$。

对于特征值 $\lambda_3=-2$ 有

$$(-2I-A)t_3=\theta$$

即当 $t_3=(t_{13},t_{23},t_{33})^T$ 时有

$$\begin{bmatrix} 0 & -2 & 1 \\ 0 & 0 & 0 \\ -1 & 4 & -2 \end{bmatrix}\begin{bmatrix} t_{13} \\ t_{23} \\ t_{33} \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

解得 $t_{33}=2t_{23}$，$t_{13}=0$，故可令 $t_3=(0,1,2)^T$，于是所求相似变换矩阵为

$$T=\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

例 6 求解微分方程组

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t}=-x_1+x_2 \\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t}=-4x_1+3x_2 \\ \dfrac{\mathrm{d}x_3}{\mathrm{d}t}=x_1+2x_3 \end{cases} \tag{12}$$

解 将微分方程组(12)写成矩阵形式，得

$$\frac{\mathrm{d}x}{\mathrm{d}t}=Ax \tag{13}$$

其中

$$A=\begin{bmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix},\ x=\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix},\ \frac{\mathrm{d}x}{\mathrm{d}t}=\begin{bmatrix} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} \\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t} \\ \dfrac{\mathrm{d}x_3}{\mathrm{d}t} \end{bmatrix}$$

由于

$$\det(\lambda I-A)=\begin{vmatrix}\lambda+1 & -1 & 0\\ 4 & \lambda-3 & 0\\ -1 & 0 & \lambda-2\end{vmatrix}=(\lambda-2)(\lambda-1)^2$$

而

$$\operatorname{rank}(I-A)=\operatorname{rank}\begin{bmatrix}2 & -1 & 0\\ 4 & -2 & 0\\ -1 & 0 & -1\end{bmatrix}=2=3-1$$

故 A 的约当标准形为

$$J=\begin{bmatrix}2 & & \\ & 1 & 1\\ & & 1\end{bmatrix}$$

令 $T=(t_1,t_2,t_3)$，则由 $AT=TJ$ 得

$$A\ t_1=2t_1$$

$$A\ t_2=t_2$$

$$A\ t_3=t_2+t_3$$

解之得

$$t_1=\begin{bmatrix}0\\0\\1\end{bmatrix},\ t_2=\begin{bmatrix}-1\\-2\\1\end{bmatrix},\ t_3=\begin{bmatrix}1\\1\\0\end{bmatrix}$$

即

$$T=\begin{bmatrix}0 & -1 & 1\\ 0 & -2 & 1\\ 1 & 1 & 0\end{bmatrix}$$

令 $x=Ty$，$y=(y_1,\ y_2,\ y_3)^T$，代入(13)式得

$$\frac{dy}{dt}=(T^{-1}AT)y=Jy$$

即

$$
\begin{cases}
\dfrac{\mathrm{d}y_1}{\mathrm{d}t}=2y_1 \\[2mm]
\dfrac{\mathrm{d}y_2}{\mathrm{d}t}=y_2+y_3 \\[2mm]
\dfrac{\mathrm{d}x_3}{\mathrm{d}t}=y_3
\end{cases}
\tag{14}
$$

由此微分方程组的第一个方程与第三个方程解得

$$y_1=k_1e^{2t},\ y_3=k_3e^t$$

将 $y_3=k_3e^t$ 代入(14)式的第二个方程得

$$\frac{\mathrm{d}y_2}{\mathrm{d}t}=y_2+k_3e^t$$

解之得 $y_2=(k_2+k_3t)e^t$，于是微分方程组(12)的解为

$$
x=Ty=\begin{bmatrix}0 & -1 & 1\\ 0 & -2 & 1\\ 1 & 1 & 0\end{bmatrix}\begin{bmatrix}k_1e^{2t}\\ (k_2+k_3t)e^t\\ k_3e^t\end{bmatrix}
$$

$$
=\begin{bmatrix}(k_3-k_2-k_3t)e^t\\ (k_3-2k_2-2k_3t)e^t\\ k_1e^{2t}+(k_2+k_3t)e^t\end{bmatrix}
$$

由以上的讨论知单纯矩阵的初等因子都是一次的。对于单纯矩阵还可作另一形式的分解。

定理 5 设 λ_i，$i\in\underline{\sigma}$ 为 $A\in C^{n\times n}$的相异特征值，m_i、α_i分别为 λ_i的代数重复度与几何重复度，若 A 是单纯矩阵，则存在 $E_i\in C^{n\times n}$，$i\in\underline{\sigma}$，使得

$$a)\quad A=\sum_{i=1}^{\sigma}\lambda_iE_i \tag{15}$$

b) $E_iE_j = \begin{cases} E_i & i=j \\ O & i \neq j \end{cases} \qquad i, j \in \underline{\sigma}$

c) $\sum_{i=1}^{\sigma} E_i = I_n$

d) $E_iA = AE_i = \lambda_i E_i \quad i \in \underline{\sigma}$

e) $\text{rank} E_i = \alpha_i (= m_i)$

f) 满足以上性质的 E_i, $i \in \underline{\sigma}$ 是唯一的。

g) $E_i = \dfrac{\phi_i(A)}{\phi_i(\lambda_i)}$

其中

$$\phi(\lambda) = \prod_{i=1}^{\sigma}(\lambda - \lambda_i), \ \phi_i(\lambda) = \prod_{\substack{j=1 \\ j\neq i}}^{\sigma}(\lambda - \lambda_j) = \frac{\phi(\lambda)}{\lambda - \lambda_i}$$

$$\phi_i(A) = \prod_{\substack{j=1 \\ j\neq i}}^{\sigma}(A - \lambda_j I)$$

证 a) 设 $A = T\Lambda T^{-1}$，其中 Λ 是以 A 的特征值为对角线元的对角线矩阵，T 与 T^{-1} 按相应的特征值分块为

$$T = (T_1 T_2 \cdots T_\sigma)$$

$$T^{-1} = \begin{bmatrix} \tilde{T}_1^T \\ \tilde{T}_2^T \\ \vdots \\ \tilde{T}_\sigma^T \end{bmatrix}$$

由 T^{-1} 的结构知

$$T_i^T \tilde{T}_j = \tilde{T}_i^T T_j = \begin{cases} I_{m_i} & j = i \\ O_{m_i \times m_j} & j \neq i \end{cases}$$

于是

$$A = T\Lambda T^{-1} = (T_1 T_2 \cdots T_\sigma)\begin{bmatrix} \lambda_1 \tilde{T}_1{}^T \\ \lambda_2 \tilde{T}_2{}^T \\ \vdots \\ \lambda_\sigma \tilde{T}_\sigma{}^T \end{bmatrix}$$

$$= \sum_{i=1}^{\sigma} \lambda_i T_i \tilde{T}_i{}^T = \sum_{i=1}^{\sigma} \lambda_i E_i$$

其中 $E_i = T_i \tilde{T}_i{}^T$, $i \in \underline{\sigma}$。

(15) 式称为单纯矩阵 A 的**谱分解**，E_i，$i \in \underline{\sigma}$ 称为 A 的**谱族**；

b) $E_i E_j = \begin{cases} T_i \tilde{T}_i{}^T T_i \tilde{T}_i{}^T = T_i I_{m_i} \tilde{T}_i{}^T = T_i \tilde{T}_i{}^T = E_i & i=j \\ T_i \tilde{T}_i{}^T T_j \tilde{T}_j{}^T = T_i O \tilde{T}_j{}^T = O & i \neq j \end{cases}$

c) $\sum_{i=1}^{\sigma} E_i = \sum_{i=1}^{\sigma} T_i \tilde{T}_i{}^T = TT^{-1} = I_n$

d) $E_i A = E_i \sum_{j=1}^{\sigma} \lambda_j E_j = \lambda_i E_i = AE_i$

e) 由 $E_i = T_i \tilde{T}_i{}^T$ 知

$\operatorname{rank} E_i \leqslant \operatorname{rank} T_i = \alpha_i$

又

$$n = \sum_{i=1}^{\sigma} \alpha_i = \operatorname{rank} I_n \leqslant \sum_{i=1}^{\sigma} \operatorname{rank} E_i$$

故

$$\alpha_i \leqslant \operatorname{rank} E_i$$

于是 $\alpha_i = \operatorname{rank} E_i$, $i \in \underline{\sigma}$

f) 若另有 $\tilde{E}_i$, $i \in \underline{\sigma}$ 也满足以上要求，故由

$$O = E_i(A\tilde{E}_j) = (E_i A)\tilde{E}_j = (\lambda_j - \lambda_i)E_i \tilde{E}_j$$

可知，当 $i \neq j$ 时，$E_i \tilde{E}_j = O$，于是

$$E_i = E_i I_n = E_i(\sum_{j=1}^{\sigma} \tilde{E}_j) = E_i \tilde{E}_i$$

$$= (\sum_{j=1}^{\sigma} E_j)\tilde{E}_i = I_n \tilde{E}_i = \tilde{E}_i \quad i \in \underline{\sigma}$$

即 A 的谱族唯一。

g) 这里给出 E_i 的另一种表示式，设

$$G_j = \frac{\phi_j(A)}{\phi_j(\lambda_j)}$$

于是

$$E_i G_j = E_i \frac{\phi_j(A)}{\phi_j(\lambda_j)} = E_i \frac{\prod\limits_{\substack{l=1\\l\neq j}}^{\sigma}(A - \lambda_l I)}{\phi_j(\lambda_j)}$$

$$= \frac{\sum\limits_{\substack{l=1\\l\neq j}}^{\sigma}(\lambda_i - \lambda_l)}{\phi_j(\lambda_j)} E_i = \begin{cases} E_i & i = j \\ O & i \neq j \end{cases}$$

即

$$E_i = \frac{\phi_i(A)}{\phi_i(\lambda_i)}$$

推论 设 $f(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_m\lambda^m$，$A \in C^{n\times n}$ 是单纯矩阵，且 A 的谱分解为 $A = \sum\limits_{i=1}^{\sigma}\lambda_i E_i$，则

$$f(A) = \sum_{i=1}^{\sigma} f(\lambda_i) E_i$$

除了单纯矩阵能与对角线阵相似之外，还有其它矩阵也具有这样的性质吗？下面继续讨论。

§3.3 正规阵及其分解

定义 设 $A \in C^{n\times n}(R^{n\times n})$，若

$$A^H A = AA^H \quad (A^T A = AA^T) \tag{1}$$

则称 A 为**复(实)正规阵**。

显然，埃尔米特矩阵($A = A^H$)、斜埃尔米特阵($A = -A^H$)、酉阵($A^{-1} = A^H$)都是复正规阵；对称矩阵($A = A^T$)、斜对称矩阵($A = -A^T$)、正交矩阵($A^{-1} = A^T$)都是实正规阵。但复正规阵不一定是埃尔米特矩阵。

例1 设

$$A = \begin{pmatrix} 1 & 1-2i \\ 2+i & 1 \end{pmatrix}$$

显然

$$AA^H = A^H A = \begin{pmatrix} 6 & 3-3i \\ 3+3i & 6 \end{pmatrix}$$

故 A 是正规阵，但 A 显然不是埃尔米特阵。

定理1 设 $A \in C^{n\times n}$，则 A 为正规阵的充要条件为：存在 $U \in U^{n\times n}$，使

$$A = U\Lambda U^H$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 的对角线元 $\lambda_i(i \in \underline{n})$ 是 A 的特征值。

证 必要性：设 A 是正规阵，由司楚尔定理有

$$A = URU^H$$

其中 U 为酉阵，R 为上三角阵，且其对角线元为 A 的特征值，于是

$$A^H = UR^H U^H$$

由于 $AA^H = A^H A$，有

$$RR^H = R^H R$$

设 $R = (r_{ij})_{n\times n}$，$RR^H = (r'_{ij})_{n\times n}$，$R^H R = (r''_{ij})_{n\times n}$，则

$$r'_{ii} = \sum_{k=1}^{n} |r_{ik}|^2 = \sum_{k=1}^{i-1} |r_{ik}|^2 + \sum_{k=i}^{n} |r_{ik}|^2 = \sum_{k=i}^{n} |r_{ik}|^2$$

$$r_{ii}'' = \sum_{k=1}^{n} |r_{ki}|^2 = \sum_{k=1}^{i} |r_{ki}|^2 + \sum_{k=i+1}^{n} |r_{ki}|^2 = \sum_{k=1}^{i} |r_{ki}|^2$$

但 $r'_{ii} = r_{ii}''$，$i \in \underline{n}$，故必有 $r_{ik} = 0 (i \neq k)$，即 R 是对角阵，且其对角线元为 A 的特征值；

充分性：设 $A = U\Lambda U^H$，U 为酉阵，$\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$，$\lambda_i (i \in \underline{n})$ 为 A 的特征值，则

$$AA^H = U\Lambda U^H U\Lambda^H U^H = U\Lambda\Lambda^H U^H$$
$$= U\Lambda^H \Lambda U^H = U\Lambda^H U^H U\Lambda U^H = A^H A$$

即 A 是正规阵。

仿单纯矩阵，正规阵也可作如下的谱分解。

定理 2 设 $A \in C^{n\times n}$，$\lambda_i (i \in \underline{\sigma})$ 为 A 的相异特征值，则 A 是正规阵的充要条件为存在 σ 个正交投影 $E_i (i \in \underline{\sigma})$，使得

a) $E_i E_j = E_i \delta_{ij}$，$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$，$i, j \in \underline{\sigma}$

b) $\sum_{i=1}^{\sigma} E_i = I_n$

c) $A = \sum_{i=1}^{\sigma} \lambda_i E_i$

证 必要性：设 A 为正规阵，由定理 1 知

$$A = U\Lambda U^H$$

将 U 按 A 的特征值分块为

$$U = (U_1 U_2 \cdots U_\sigma)$$

于是 $U_i^H U_i = I_{m_i}$，$AU_i = \lambda_i U_i (i \in \underline{\sigma})$，令 $E_i = U_i U_i^H$，易知

$$E_i = E_i^H = E_i^2 \qquad i \in \underline{\sigma}$$

即 E_i，$i \in \underline{\sigma}$ 是正交投影，且

a) $E_i E_j = U_i U_i^H U_j U_j^H = E_i \delta_{ij}$

b) $I_n = UU^H = \sum_{i=1}^{\sigma} U_i U_i^H = \sum_{i=1}^{\sigma} E_i$

c)　$A = U\Lambda U^H = \sum_{i=1}^{\sigma} \lambda_i U_i U_i^H = \sum_{i=1}^{\sigma} \lambda_i E_i$

充分性：设有 $E_i = E_i^H = E_i^2 (i \in \underline{\sigma})$，且满足定理中的 a)、b)、c)，于是

$$AA^H = \left(\sum_{i=1}^{\sigma} \lambda_i E_i\right)\left(\sum_{j=1}^{\sigma} \overline{\lambda_j} E_j^H\right)$$

$$= \sum_{i=1}^{\sigma} \lambda_i \overline{\lambda_i} E_i = \sum_{i=1}^{\sigma} \overline{\lambda_i} \lambda_i E_i = A^H A$$

即 A 是正规阵。

从以上两个定理看出，正规阵必为单纯阵，且还有以下性质。

定理 3　设 $A \in C^{n \times n}$ 是正规阵，则

a)　对于 A 与 A^H 存在同一酉阵 U，使 $U^H A U$、$U^H A^H U$ 均为对角矩阵；

b)　A 是单纯矩阵；

c)　若 $Ax = \lambda_i x$，则 $A^H x = \overline{\lambda_i} x$，$x \neq \theta$；

d)　若 $Ax = \lambda x$，$Ay = \mu y$，$x \neq \theta$，$y \neq \theta$，$\lambda \neq \mu$，则

$$(x, y) = 0$$

证　a)　因为 A 是正规阵，故存在酉阵 U，使得

$$A = U\Lambda U^H$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 的对角线元 $\lambda_i (i \in \underline{n})$ 是 A 的特征值，由此可得

$$\Lambda^H = U^H A^H U$$

显然，$\Lambda^H = \mathrm{diag}(\overline{\lambda_1}, \overline{\lambda_2}, \cdots, \overline{\lambda_n})$ 的对角线元 $\overline{\lambda_i} (i \in \underline{n})$ 是 A^H 的特征值；

b)　因为此时 A 相似于以 A 的特征值为对角线元的对角阵，故 A 是单纯阵；

c） 设 $U=(U_1U_2\cdots U_n)$，于是 $A\,U_i=\lambda_iU_i$，由 a）有 $A^HU_i=\overline{\lambda}_iU_i$，故当 $Ax=\lambda_ix$ 时，$A^Hx=\overline{\lambda}_ix$，$x\neq\theta$；

d） 由

$$(Ax,y)=(\lambda x,\ y)=\overline{\lambda}(x,y)$$
$$(Ax,y)=(x,A^Hy)=(x,\overline{\mu}y)=\overline{\mu}(x,y)$$

得

$$(\overline{\lambda}-\overline{\mu})(x,y)=0$$

由于 $\lambda\neq\mu$，故 $\overline{\lambda}\neq\overline{\mu}$，所以 $(x,y)=0$。

正规阵的两种分解形式在结构上与单纯矩阵的两种分解形式相似，但有差异。单纯矩阵与对角阵相似，而正规阵与对角阵酉相似。其次，正规阵的谱分解中，谱族是正交投影，单纯矩阵的谱族不一定具有这一性质。正规阵的特征子空间还是正交子空间。

§3.4 埃尔米特矩阵及其分解

埃尔米特矩阵是正规矩阵，因此它具有正规矩阵所具有的两种形式的分解式，但两种形式也有差异，这种差异将在以下对埃尔米特矩阵的性质讨论中得知。埃尔米特矩阵是酉空间给定的内积在选定基下的矩阵，也是复二次型的系数矩阵，因此应用十分广泛。下面讨论它的性质及复二次型。

定理 1 设 $A=A^H\in C^{n\times n}$，则

a） $(Ax,\ y)=(x,\ Ay)\qquad \forall\ x,y\in C^n$

b） A 的特征值均为实数。

证 a） 显然。

b） 设 $\theta\neq x\in C^n$，且 $Ax=\lambda x$，则

$$(Ax,x)=(\lambda x,x)=\overline{\lambda}(x,x)$$

$$(Ax,x)=(x,Ax)=\lambda(x,x)$$

于是

$$(\bar{\lambda}-\lambda)(x,x)=0$$

由于 $x\neq\theta$，故 $(x,x)>0$，所以 $\lambda-\bar{\lambda}=0$，即 $\lambda\in R$。

推论 若 $A=A^H\in C^{n\times n}$，且 $\text{rank}A=r$，则 A 与矩阵

$$\begin{bmatrix} I_p & & \\ & -I_{r-p} & \\ & & O \end{bmatrix}$$

合同。

证 由于 A 是正规阵，故存在酉阵 U 及对角阵 $\Lambda=\text{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，$\lambda_i\in R$，$i\in \underline{n}$，使得

$$A=U\Lambda U^H$$

又由于 $\text{rank}A=\text{rank}\Lambda=r$，故 Λ 的对角线元只能有 r 个不为零。因为它们都是实数，不妨设其中 $\lambda_j>0$，$j\in \underline{p}$，λ_{p+1}、λ_{p+2}、$\cdots$、λ_r 均为负数，$\lambda_{r+1}=\lambda_{r+1}=\cdots=\lambda_n=0$。令

$$\mu_k=\sqrt{|\lambda_k|}\qquad k\in \underline{r}$$

$$M=\text{diag}(\mu_1,\mu_2,\cdots,\mu_r,1,1,\cdots,1)$$

于是

$$\Lambda=M\begin{bmatrix} I_p & & \\ & -I_{r-p} & \\ & & O \end{bmatrix}M^H$$

再令 $T=UM$，则

$$A=T\begin{bmatrix} I_p & & \\ & -I_{r-p} & \\ & & O \end{bmatrix}T^H$$

这里数 p 由 A 唯一确定，也称为 A 的**正惯性指数**，而 $q=r-p$ 称为 A 的**负惯性指数**。它们均为合同变换下的不变量。

例 1 若 $A=-A^H\in C^{n\times n}$，则 A 的特征值为纯虚数。

证 设 $Ax=\lambda x$，$x\neq\theta$，则由

$$(Ax,x)=\bar{\lambda}(x,x)$$

$$(Ax,x)=-(A^Hx,x)=-(x,Ax)=-\lambda(x,x)$$

得

$$(\lambda+\bar{\lambda})(x,x)=0$$

由于 $(x,x)>0$，故 $\lambda+\bar{\lambda}=0$，即 $\mathrm{Re}(\lambda)=0$，所以 λ 为纯虚数。

在复二次型(也可称为埃尔米特型)中，正定二次型应用最为广泛，它可用来定义内积和范数，也可用来刻画力学中的能量以及系统中的品质指数。

定义 1 若 $A=A^H\in C^{n\times n}$，对任意的 $\theta\neq x\in C^n$，都有

$$f(x)=x^HAx>0(\geqslant 0)$$

则称二次型 $f(x)$ 是**正定(半正定)**的。

正定(半正定)二次型的系数矩阵 A 称为正定(半正定)矩阵，记为 $A>O(\geqslant O)$，因此，复二次型正定性的讨论，可变为对其系数矩阵 A 的正定性的讨论。

定理 2 设 $A=A^H\in C^{n\times n}$，则下列条件等价：

a) A 正定；

b) A 的特征值全为正实数；

c) A 与单位矩阵 I_n 合同；

d) A 的顺序主子式全为正。

证 $a)\Rightarrow b)$。因为 $A=A^H$，故存在酉阵 U，使 $U^HAU=\Lambda$，在 $f(x)=x^HAx$ 中，令 $x=Uy$，得

$$f(x)=y^H\Lambda y=\sum_{i=1}^{n}\lambda_i|y_i|^2>0$$

其中 $y=(y_1, y_2, \cdots, y_n)^T\neq\theta$，故 $\lambda_i>0$，$i\in\underline{n}$；

$b)\Rightarrow a)$ 显然；

$b)\Leftarrow\Rightarrow c)$ 显然；

$a)\Rightarrow d)$ 记 $F_k=(e_1, e_2, \cdots, e_k)$，其中 $e_i=(0, 0, \cdots, \underset{\text{第}i\text{个}}{0, 1}, 0, \cdots, 0)^T$ $i\in\underline{k}$ 为 C^n 的单位向量，取 $t\in C^k$，则

$$(F_kt)^HA(F_kt)=t^HA_kt>0$$

其中 A_k 为 A 的 k 阶顺序主子矩阵，显然，$\det A_k>0$，其中 $\det A_k$ 即 A 的 k 阶顺序主子式；

$d)\Rightarrow a)$ 记 $\Delta_k=\det A_k$，于是 A 可唯一地分解为

$$A=\tilde{L}D\tilde{U}$$

其中 $\tilde{L}$、$\tilde{U}$ 分别为单位下三角阵与单位上三角阵，$D=\operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)=\operatorname{diag}\left(\Delta_1, \frac{\Delta_2}{\Delta_1}, \cdots, \frac{\Delta_n}{\Delta_{n-1}}\right)$。由于 $A=A^H$，故

$$A=\tilde{U}^HD^H\tilde{L}^H$$

由 A 分解式的唯一性得 $\tilde{L}=\tilde{U}^H$。若令

$$D^{\frac{1}{2}}=\operatorname{diag}\left(\sqrt{\Delta_1}, \sqrt{\frac{\Delta_2}{\Delta_1}}, \cdots, \sqrt{\frac{\Delta_n}{\Delta_{n-1}}}\right)$$

则

$$A=(\tilde{L}D^{\frac{1}{2}})(\tilde{L}D^{\frac{1}{2}})^H=T^HT$$

其中 $T=(\tilde{L}D^{\frac{1}{2}})^H$，即 $c)$ 成立，从而 $a)$ 成立。

在定理的证明中得知，正定埃尔米特矩阵 A，可唯一地分解为

$$A=LL^H$$

其中 $L=\tilde{L}D^{\frac{1}{2}}$ 为下三角阵，正定矩阵的这种分解称为楚来斯基(Cholesky)分解。

下面继续讨论正定埃尔米特矩阵的性质。

定理 3 若 $A=A^H\in C^{n\times n}$ 正定，则

a) A 的对角线元均大于零；

b) $A=B^2$，B 是正定矩阵；

c) A 的任意 k 行与对应的 k 列组成的主子阵

$$A\begin{pmatrix} i_1 i_2\cdots i_k \\ i_1 i_2\cdots i_k \end{pmatrix}=\begin{bmatrix} \alpha_{i_1i_1} & \alpha_{i_1i_2} & \cdots & \alpha_{i_1i_k} \\ \alpha_{i_2i_1} & \alpha_{i_2i_2} & \cdots & \alpha_{i_2i_k} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{i_ki_1} & \alpha_{i_ki_2} & \cdots & \alpha_{i_ki_k} \end{bmatrix}$$

正定。

其中 $1\leqslant i_1<i_2<\cdots<i_k\leqslant n$。

证 a) 由 A 正定，在 $f(x)=x^HAx$ 中，取 $x=e_i=(0,\cdots,0,\underset{\text{第}i\text{个}}{1},0,\cdots,0)^T$，则

$$e_i^HAe_i=\alpha_{ii}>0 \qquad i\in \underline{n}$$

b) 设 $A=U\Lambda U^H$，U 为酉阵，$\Lambda=\operatorname{diag}(\lambda_1,\lambda_2,\cdots\lambda_n)$，$\lambda_i>0(i\in\underline{n})$，记

$$\Lambda^{\frac{1}{2}}=\operatorname{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots\sqrt{\lambda_n})$$

则

$$\begin{aligned} A&=U\Lambda U^H=U\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}U^H \\ &=(U\Lambda^{\frac{1}{2}}U^H)(U\Lambda^{\frac{1}{2}}U^H)=(U\Lambda^{\frac{1}{2}}U^H)^2=B^2 \end{aligned}$$

其中 $B=U\Lambda^{\frac{1}{2}}U^H$ 显然正定。

c) 令 $\tilde{F}_k=(e_{i_1},e_{i_2},\cdots,e_{i_k})$，$1\leqslant i_1<i_2<\cdots<i_k\leqslant n$，其中

$e_{i_r}(r\in \underline{k})$是第 i_r 个分量为 1，其余分量为 0 的 n 维单位向量，取 $t\in C^k$，得

$$(\widetilde{F}_k t)^H A(\widetilde{F}_k t)=t^H A\begin{pmatrix} i_1 i_2\cdots i_k \\ i_1 i_2\cdots i_k \end{pmatrix} t>0$$

故 $A\begin{pmatrix} i_1 i_2\cdots i_k \\ i_1 i_2\cdots i_k \end{pmatrix}$正定。

例 2 若 $A=(a_{ij})_{n\times n}\in C^{n\times n}$是正定埃尔米特矩阵，则

$$\prod_{i=1}^{n} a_{ii} > \det A$$

证 由正定矩阵的楚来斯基分解知，存在下三角矩阵 L，使得

$$A=LL^H$$

若记 $L=(l_{ij})_{n\times n}$，则

$$a_{11}=|l_{11}|^2$$

$$a_{22}=|l_{21}|^2+|l_{22}|^2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_{nn}=|l_{n1}|^2+|l_{n2}|^2+\cdots+|l_{nn}|^2$$

于是

$$\prod_{i=1}^{n} a_{ii} > \prod_{i=1}^{n}|l_{ii}|^2=\det A$$

半正定矩阵应用也很广泛，它有与正定矩阵相应的性质。

定理 4 设 $A=A^H\in C^{n\times n}$，则下列条件等价：

a) A 半正定；

b) A 的特征值非负；

c) A 与$\begin{pmatrix} I_r & \\ & O \end{pmatrix}$合同，$r=\text{rank}A$；

d) A 的所有主子式皆非负，即

$$\det A\begin{pmatrix} i_1 i_2 \cdots i_k \\ i_1 i_2 \cdots i_k \end{pmatrix} \geqslant 0 \quad 1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$$

定理 5 设 $A \in A^H \in C^{n\times n}$ 半正定，则

a) A 的对角线元均非负；

b) $A = B^2$，B 是半正定矩阵。

定理 4、定理 5 的证明与定理 2、定理 3 的证明类似。

任意一个二次型均可化为标准型，即系数矩阵的对角化，在实际问题中，有时需要考虑两个埃尔米特矩阵的同时对角化。例如，广义特征值问题：

$$Bx = \lambda Ax$$

就是常在 $A > O$ 时，用 A、B 同时对角化来讨论的。

定理 6 若 $A = A^H \in C^{n\times n}$ 正定，$B = B^H \in C^{n\times n}$，则存在满秩矩阵 $T \in C^{n\times n}$，使

$$T^H A T = I_n \tag{1}$$

$$T^H B T = \Lambda \tag{2}$$

其中 Λ 是对角线矩阵。

证 由 A 正定知，存在满秩矩阵 $P \in C^{n\times n}$，使得 $A = P^H P$，即

$$(P^H)^{-1} A P^{-1} = I_n \tag{3}$$

令 $\tilde{B} = (P^H)^{-1} B P^{-1}$，由 $B = B^H$ 知 $\tilde{B} = \widetilde{B^H}$，于是存在 $U \in U^{n\times n}$，使

$$U^H \tilde{B} U = \Lambda \tag{4}$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$，$\lambda_i (i \in n)$ 是 $\tilde{B}$ 的特征值，再令 $T = P^{-1} U$，则由(3)、(4)即知(1)、(2)式成立。

注意 由于 $\det A>0$，又

$$\det(\lambda I-\widetilde{B})=\det[\lambda(P^H)^{-1}AP^{-1}-(P^H)^{-1}BP^{-1}]$$
$$=\det(P^H)^{-1}(\lambda A-B)P^{-1}=\det(P^H)^{-1}A(\lambda I-A^{-1}B)P^{-1}$$
$$=\det(P^H)^{-1}\det P^{-1}\det A\det(\lambda I-A^{-1}B)=\det(\lambda I-A^{-1}B)$$

故 $\widetilde{B}$ 的特征值就是 $A^{-1}B$ 的特征值。但 $A^{-1}B$ 一般不再是埃尔米特阵，而采用同时对角化，则可使 $Bx=\lambda Ax$ 在 A、B 均为埃尔米特阵且 $A>O$ 时，有

$$(T^HBT)x=\lambda x(\text{或 } \Lambda x=\lambda x)$$

显然 T^HBT 仍为埃尔米特矩阵，这就将广义特征值问题转化为一般的埃尔米特矩阵的特征值问题了。

例 3 把矩阵

$$A=\begin{pmatrix}1 & i\\ -i & 2\end{pmatrix},\quad B=\begin{pmatrix}1 & 1+i\\ 1-i & 1\end{pmatrix}$$

同时对角化。

解 易知 A 正定，且 $\det A=1$

$$A^{-1}=\begin{pmatrix}2 & -i\\ i & 1\end{pmatrix}$$

$$A^{-1}B=\begin{pmatrix}1-i & 2+i\\ 1 & i\end{pmatrix}$$

$$\det(\lambda I-A^{-1}B)=\lambda^2-\lambda-1=\left(\lambda-\frac{1+\sqrt{5}}{2}\right)\left(\lambda-\frac{1-\sqrt{5}}{2}\right)$$

故 A 与 I_2 合同，B 合同于

$$\begin{bmatrix}\frac{1+\sqrt{5}}{2} & 0\\ 0 & \frac{1-\sqrt{5}}{2}\end{bmatrix}$$

推论 若 A、$B\in C^{n\times n}$ 均为正定埃尔米特矩阵，且 $A-B$

$\geqslant O$（$A-B$ 半正定也可记为 $A\geqslant B$），则

$$B^{-1}\geqslant A^{-1}>O$$

证 由定理 6 知，存在可逆矩阵 T，使

$$T^H AT=\Lambda=\text{diag}(\mu_1,\mu_2,\cdots,\mu_n) \qquad (\mu_i>0,\ i\in \underline{n})$$

$$T^H BT=I_n$$

从而有

$$T^{-1}A^{-1}(T^H)^{-1}=\Lambda^{-1}=\text{diag}\left(\frac{1}{\mu_1},\frac{1}{\mu_2},\cdots,\frac{1}{\mu_n}\right)$$

$$T^{-1}B^{-1}(T^H)^{-1}=I_n$$

由于 $A\geqslant B$，故 $\Lambda\geqslant I$，所以 $\mu_i\geqslant 1(i\in \underline{n})$，于是

$$1\geqslant\frac{1}{\mu_i}>0 \qquad i\in \underline{n}$$

即

$$I\geqslant\Lambda^{-1}>0$$

所以

$$B^{-1}\geqslant A^{-1}>O$$

在实际问题中，常需要在一定条件限制下讨论二次型的条件极值。利用瑞利(Rayleigh)商是讨论二次型条件极值的一条途径，现对瑞利商作简要介绍。

定义 2 设 $A=A^H\in C^{n\times n}$，则映射

$$R_A(x)=\frac{x^H Ax}{x^H x} \qquad \forall\ \theta\neq x\in C^n$$

称为矩阵 A 的**瑞利商**。

由于 $x^H Ax=(x,\ Ax)=(Ax,\ x)=(\overline{Ax,\ x})$，故瑞利商 $R_A(x)$ 是映射 $C^n-\{\theta\}\rightarrow R$。

例 4 设

$$A=\begin{pmatrix}1 & i\\ -i & 1\end{pmatrix} \quad \theta\neq x=\begin{pmatrix}x_1\\ x_2\end{pmatrix}$$

则

$$R_A(x)=\frac{x^H A x}{x^H x}=\frac{|x_1|^2+|x_2|^2+i(\bar{x}_1 x_2-x_1\bar{x}_2)}{|x_1|^2+|x_2|^2}$$

定理 7 设 $R_A(x)$是 $A=A^H\in C^{n\times n}$的瑞利商，则

a) $R_A(\lambda x)=R_A(x) \qquad \forall\ 0\neq\lambda\in C, \forall\ \theta\neq x\in C^n$

b) $\min\limits_i \lambda_i(A)\leqslant R_A(x)\leqslant\max\limits_i\lambda_i(A)$

c) 存在 $\theta\neq x_i\in C^n$，$i\in\underline{n}$，使得

$$R_A(x_i)=\lambda_i(A)$$

d) 若 $V(A)\subset R$ 为 $R_A(x)$的值域，$U\in U^{n\times n}$，则

$$V(A)=V(U^H AU)$$

证 a) 显然；

b) 设 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$为 A 的特征值，x_1、x_2、$\cdots$、x_n 为对应的特征向量，它们构成 C^n 的标准正交基。记 $U=(x_1\ x_2\cdots\ x_n)$，则

$$U^H AU=\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

令 $x=Uy$，$\theta\neq y\in C^n$，则

$$R_A(x)=\frac{x^H Ax}{x^H x}=\frac{y^H U(^H AU)y}{y^H y}=\frac{y^H\Lambda y}{y^H y}$$

$$=\frac{\lambda_1 y_1{}^2+\lambda_2 y_2{}^2+\cdots+\lambda_n y_n{}^2}{y_1{}^2+y_2{}^2+\cdots+y_n{}^2}\leqslant\lambda_1=\max_i\lambda_i(A)$$

同理可证

$$R_A(x)\geqslant\lambda_n=\min_i\lambda_i(A)$$

c) $R_A(x_i)=\dfrac{x_i{}^H Ax_i}{x_i{}^H x_i}=\dfrac{\lambda_i x_i{}^H x_i}{x_i{}^H x_i}=\lambda_i \quad i\in\underline{n}$

d) 任取 $a\in V(A)$，存在 $\theta\neq x_0\in C^n$，使得

$$a=R_A(x_0)=\frac{x_0{}^H Ax_0}{x_0{}^H x_0}$$

显然存在 $U\in U^{n\times n}$及 y_0，使得 $x_0=Uy_0$，于是

$$a=\frac{y_0^H U^H A U y_0}{y_0^H y_0}\in V(U^H AU)$$

即 $V(A)\subset V(U^H AU)$。反之，可证 $V(U^H AU)\subset V(A)$，于是 d)成立。

作为本节的结束，简要介绍一下"广义正定矩阵"的概念及部分新结果。

关于矩阵的正定性研究，过去只限于对称矩阵和埃尔米特矩阵，近年来已突破了这一限制，从而获得了广义正定矩阵的概念。

定义 3 设 $A\in R^{n\times n}$，若对任何 $\theta\neq x\in R^n$，都有

$$x^T Ax>0$$

则称 A 为**正定矩阵**。

显然，如果 A 正定，则 A^T 也正定；如果 A、$B\in R^{n\times n}$均正定，则 $A+B$ 也正定。

记 $A=S+K$，其中 $S=\frac{1}{2}(A+A^T)$，$K=\frac{1}{2}(A-A^T)$，并分别称 S、K 为 A 的**对称分量**及**斜对称分量**。

定理 8 设 $A\in R^{n\times n}$，则

a) A 正定的充要条件为其对称分量 S 正定。

b) A 正定的充要条件为存在 n 阶可逆矩阵 P，使得

$$P^T AP=\mathrm{diag}\left(\begin{pmatrix}1 & a_1\\ -a_1 & 1\end{pmatrix},\cdots,\begin{pmatrix}1 & a_s\\ -a_s & 1\end{pmatrix},1,\cdots,1\right)$$

其中 $a_1\geqslant a_2\geqslant\cdots\geqslant a_s>0$。

定理 9 设 $A=S+K\in R^{n\times n}$，其中 S、K 为 A 的对称分量与斜对称分量，则

a) $\min\{\lambda(S)\}\leqslant\mathrm{Re}\{\lambda(A)\}\leqslant\max\{\lambda(S)\}$

b) 当 A 正定时，有

$$\det A\geqslant\det S+\det K$$

c) 当 A 正定时，$\det A>0$；

d) 当 A 正定时，

$$\det A\begin{pmatrix}1,2,\cdots,j\\1,2,\cdots,j\end{pmatrix}\geqslant\det S\begin{pmatrix}1,2,\cdots,j\\1,2,\cdots,j\end{pmatrix}+\det K\begin{pmatrix}1,2,\cdots,j\\1,2,\cdots,j\end{pmatrix},j\in\underline{n}$$

其中 $\det B\begin{pmatrix}1,2,\cdots,j\\1,2,\cdots,j\end{pmatrix}$表示矩阵 B 的 j 阶顺序主子式。

e) 当 A 正定时，A^{-1}也是正定的。

定理 8、定理 9 的证明请看《数学的认识与实践》杂志1985 年第三期，李炯生所发表的〈实方阵的正定性〉一文。

定义 4 设 $A\in R^{n\times n}$，若对任何 $\theta\neq x\in R^n$，都有正对角矩阵 $D=D_x>O$，使

$$x^TD_xAx>0$$

则称 A 为正定矩阵。

记对称正定矩阵类为 P_S；满足定义 3 的正定矩阵类记为 P_L；满足定义 4 的正定矩阵类，当 D_x 与 x 无关时记为 P_D；否则记为 P_B。

定理 10 $P_S\subset P_L\subset P_D\subset P_B$。

定理 11 设 $A\in R^{n\times n}$，则下列条件等价：

a) $A\in P_B$

b) A 的各阶主子式全为正；

c) 对任意的 $\theta\neq x\in R^n$，取 $y=Ax$，则必有 $k(1\leqslant k\leqslant n)$，使 $x_ky_k>0$，其中 x_k、y_k 分别为 x 与 y 的第 k 个分量；

d) 对任何的 $\theta\neq x\in R^n$，都有非负对角阵 $H_x\geqslant O$，使 $x^HH_xAx\geqslant 0$；

e) A 的每一个主子矩阵的实特征值为正；

f) 对任意的 $\theta\neq x\in R^n$，都有正对角矩阵 $D=D_x>O$，使得 $x^TA^TD_xx>0$；

$g)$ 对任意的 $\theta \neq x \in R^n$，都有非负对角矩阵 $H_x \geq O$，使得 $x^T A^T H_x x > 0$；

$h)$ $A^T \in P_D$

$i)$ 对任何 n 阶正对角矩阵 D_1 与 D_2，都有 $D_1 A D_2 \in P_D$；

$j)$ 对任何的 $\theta \neq x \in R^n$，都有正对角矩阵 $D = D_x > O_{k \times k}$，使

$$x^T D_x A\begin{pmatrix} i_1 i_2 \cdots i_k \\ i_1 i_2 \cdots i_k \end{pmatrix} x > 0 \qquad k \in \underline{n}$$

定理 12 设 $A \in R^{n \times n}$，D 为 n 阶正对角矩阵，则下列条件等价：

$a)$ $A \in P_D$

$b)$ 对任意一个 n 阶正对角阵 D_1，$D_1 A \in P_{DD_1^{-1}}$；

$c)$ $A^T D \in P_L$

$d)$ $DA \in P_L$

$e)$ $D^{-1} A \in P_L$

$f)$ $AD^{-1} \in P_L$

$g)$ $A^T \in P_{D^{-1}}$

$h)$ 对任意的 n 阶正对角矩阵 D_1 与 D_2，均有

$$D_1 A D_2 \in P_{D_2 D D_1^{-1}}$$

$i)$ 对任意的正对角矩阵 D_1，$AD \in P_{D_1 D}$；

$j)$ 对 A 的任意主子阵 A_1，都有正对角阵 $D_1 > O$，使得 $A_1 \in P_{D_1}$。

定理 10、定理 11 及定理 12 的证明，请看《数学学报》杂志 1985 年 11 月，佟文廷发表的〈 广义正定矩阵 〉一文。

以上内容是自 1970 年 C. R. Johnson 推广正定矩阵概念之后，我国数学工作者取得的新成果。

§3.5 矩阵的最大秩分解

以上各节介绍了 n 阶方阵的几种分解，从本节开始，将介绍几种常用的长方阵的分解。

本节给出将矩阵 $A\in C^{m\times n}$ 分解为两个与 A 同秩的因子的积的具体方法，并讨论不同分解之间的关系。它们在广义逆矩阵的讨论中，将是十分重要的。

定理 1 设 $A\in C_r^{m\times n}$①，则存在矩阵 $B\in C_r^{m\times r}$、$C\in C_r^{r\times n}$，使得

$$A=BC$$

证 设 $A=(A_{11},\ A_{12})P$，其中 $A_{11}\in C_r^{m\times r}$，它由 A 的 r 个线性无关列所组成，A_{12}为 A 的其余 $n-r$ 列所组成的矩阵，$P\in C_n^{n\times n}$ 为初等列变换矩阵之积。由于 A_{12}的列均为 A_{11}的列的线性组合，故存在矩阵 $D\in C^{r\times(n-r)}$，使得

$$A_{12}=A_{11}D$$

于是

$$A=(A_{11},\ A_{11}D)P=A_{11}(I_r,\ D)P$$

令

$$B=A_{11},\ C=(I_r,\ D)P$$

显然有 $B\in C_r^{m\times r}$，$C\in C_r^{r\times n}$，且 $A=BC$。

矩阵的这种分解，称为**最大秩分解(满秩分解)**。

定理的证明过程给出求 B、C 的方法，可归纳如下：

将 A 进行行初等变换，化为行标准形，即将 A 变为如下形式的矩阵

---

① 记号 $C_r^{m\times n}$ 表示矩阵集合 $C^{m\times n}$中秩为 r 的矩阵所构成的子集。

$$\widetilde{A}_r=\begin{bmatrix} 0\cdots 0 & 1 & *\cdots * & 0 & *\cdots * & 0 & *\cdots * \\ 0\cdots 0 & 0 & 0\cdots 0 & 1 & *\cdots * & 0 & *\cdots * \\ \vdots & & & & & \vdots & \vdots \\ & & & & & 0 & \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & *\cdots * \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

（上方标注：k_1列、k_2列、k_r列；右侧标注：r个元素不全为零的行；$n-r$个元素全为零的行）

其中“$*$”表示不一定为 0 的元素，在 $\widetilde{A}_r$中第 k_j 列的元素除第 j 个元素为 1 外，其余元素均为 $0(j\in \underline{r})$。于是 A 中 k_1、k_2、$\cdots$、k_r 列的元素组成的 $m\times r$ 阶矩阵就是 B。而在 $\widetilde{A}_r$中除去下面的 $n-r$ 个元素全为 0 的行外，所得的 $r\times n$ 阶矩阵即为 C。

例 1 求矩阵

$$A=\begin{bmatrix} 1 & 4 & -1 & 5 & 6 \\ 2 & 0 & -4 & 0 & -14 \\ -1 & 2 & 0 & 0 & 1 \\ 1 & -2 & -1 & -1 & -6 \end{bmatrix}$$

的最大秩分解。

解 将 A 进行行初等变换，化为行标准形

$$A\to\begin{bmatrix} 1 & 0 & 0 & 2 & 3 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

于是取 A 的前三列组成矩阵

$$B=\begin{bmatrix} 1 & 4 & -1 \\ 2 & 0 & -4 \\ -1 & 2 & 0 \\ 1 & -2 & -1 \end{bmatrix}$$

再取 A 的行标准形的前三个非零行，组成矩阵

$$C=\begin{bmatrix}1 & 0 & 0 & 2 & 3\\ 0 & 1 & 0 & 1 & 2\\ 0 & 0 & 1 & 1 & 5\end{bmatrix}$$

容易验证 $A=BC$。

对于矩阵的性质，一般"行"具有的性质，"列"也具有。例如，在例 1 中，将 A 进行列变换化为列标准形得

$$A\rightarrow\begin{bmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ -\frac{1}{5} & \frac{3}{10} & -\frac{3}{5} & 0 & 0\end{bmatrix}$$

于是，将 A 的列标准形的前三列取作因子

$$\tilde{B}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ -\frac{1}{5} & \frac{3}{10} & -\frac{3}{5}\end{bmatrix}$$

再将 A 的前三行取作因子

$$\tilde{C}=\begin{bmatrix}1 & 4 & -1 & 5 & 6\\ 2 & 0 & -4 & 0 & -14\\ -1 & 2 & 0 & 0 & 1\end{bmatrix}$$

容易验证 $A=\tilde{B}\tilde{C}$，且

$$\operatorname{rank}\tilde{B}=\operatorname{rank}\tilde{C}=\operatorname{rank}A=3$$

由此可见，矩阵 A 的最大秩分解不是唯一的。但最大秩分解之间，有如下关系。

定理 2 设 $A\in C_r^{m\times n}$，且

$$A=BC=\tilde{B}\tilde{C}$$

均为 A 的最大秩分解，则

a） 存在矩阵 $Q\in C_r^{r\times r}$，使得

$$B=\tilde{B}Q,\ C=Q^{-1}\tilde{C} \tag{1}$$

b） $C^H(CC^H)^{-1}(B^HB)^{-1}B^H=\tilde{C}^H(\tilde{C}\tilde{C}^H)^{-1}(\tilde{B}^H\tilde{B})^{-1}\tilde{B}^H$ (2)

证 a） 有 $BC=\tilde{B}\tilde{C}$，有

$$BCC^H=\tilde{B}\tilde{C}C^H \tag{3}$$

又由 $\text{rank}C=\text{rank}CC^H=r$，$CC^H\in C_r^{r\times r}$，所以矩阵 CC^H 可逆。在(3)式两端同时右乘 $(CC^H)^{-1}$ 得

$$B=\tilde{B}\tilde{C}C^H(CC^H)^{-1}=\tilde{B}Q_1 \tag{4}$$

其中 $Q_1=\tilde{C}C^H(CC^H)^{-1}$。

同理可得

$$C=(B^HB)^{-1}B^H\tilde{B}\tilde{C}=Q_2\tilde{C} \tag{5}$$

其中 $Q_2=(B^HB)^{-1}B^H\tilde{B}$。

将(4)、(5)代入 $BC=\tilde{B}\tilde{C}$ 得

$$\tilde{B}\tilde{C}=\tilde{B}Q_1Q_2\tilde{C}$$

上式两端左乘 $\tilde{B}^H$、右乘 $\tilde{C}^H$ 得

$$\tilde{B}^H\tilde{B}\tilde{C}\tilde{C}^H=\tilde{B}^H\tilde{B}Q_1Q_2\tilde{C}\tilde{C}^H$$

又由于 $\tilde{B}^H\tilde{B}$、$\tilde{C}\tilde{C}^H$ 均可逆，上式两端分别左乘 $(\tilde{B}^H\tilde{B})^{-1}$，右乘 $(\tilde{C}\tilde{C}^H)^{-1}$ 得

$$I_r=Q_1Q_2$$

显然 Q_1、Q_2 均为 r 阶方阵，若记 $Q_1=Q$，则 $Q_2=Q^{-1}$，即(1)式成立；

b） 由(1)有

$$\begin{aligned}C^H(CC^H)^{-1}(B^HB)^{-1}B^H&=(Q^{-1}\tilde{C})^H[Q^{-1}\tilde{C}(Q^{-1}\tilde{C})^H]^{-1}[(\tilde{B}Q)^H\\&\times(\tilde{B}Q)]^{-1}(\tilde{B}Q)^H=\tilde{C}^H(Q^{-1})^H[Q^{-1}(\tilde{C}\tilde{C}^H)(Q^{-1})^H]^{-1}\\&\times[Q^H(\tilde{B}^H\tilde{B})Q]^{-1}Q^H\tilde{B}^H=\tilde{C}^H(Q^{-1})^H[(Q^{-1})^H]^{-1}(\tilde{C}\tilde{C}^H)^{-1}\\&\times QQ^{-1}(\tilde{B}^H\tilde{B})^{-1}(Q^H)^{-1}Q^H\tilde{B}^H=\tilde{C}^H(\tilde{C}\tilde{C}^H)^{-1}(\tilde{B}^H\tilde{B})^{-1}\tilde{B}^H\end{aligned}$$

即(2)式成立。

(2)式表明，矩阵 A 的最大秩分解虽不唯一，但由最大秩分解所作出这种形式的乘积

$$C^H(CC^H)^{-1}(B^HB)^{-1}B^H$$

是相同的，这个乘积表达式正是今后要用的 A 的广义逆矩阵中的伪逆矩阵。

§3.6 矩阵的 QR 分解

1961 年 J. G. F. Francis 提出求一般矩阵全部特征值、特征向量的最有效的方法——QR 算法，它在广义逆矩阵计算、最小二乘问题等方面都十分重要，其基础即矩阵的 QR 分解，这里只给出分解定理。

定理 设 $A\in C_r^{m\times n}$，则 A 可分解为

$$A=QR$$

其中 $Q\in C_r^{m\times r}$，且 $Q^HQ=I_r$，$R\in C_r^{r\times n}$。

矩阵 A 的这种分解称为 **QR 分解**。

证 设 $A=BC$ 为 A 的最大秩分解，对于 B 的 r 个线性无关列进行标准正交化知，存在可逆上三角阵 D，使得

$$B=QD$$

其中 $Q\in C_r^{m\times r}$，且 $Q^HQ=I_r$，$D\in C_r^{r\times r}$为上三角阵。令 $DC=R$，则得

$$A=QR$$

从上述定理可知，A的QR分解是一种特殊的最大秩分解。

推论 1 若 $A\in C_r^{m\times n}$，则 A 可唯一地分解为

$$A=QR$$

其中 $Q\in C_r^{m\times r}$，且 $Q^HQ=I_r$，$R\in C_r^{r\times r}$为具正对角元的上三

角阵。

证 将 A 的 r 个线性无关列标准正交化即得

$$A=QR$$

其中 $Q\in C_r^{m\times r}$，且 $Q^HQ=I_r$，$R\in C_r^{r\times r}$为具正对角元的上三角阵。

现再证这种分解是唯一的。设 A 有两种上述分解

$$A=Q_1R_1=Q_2R_2 \tag{1}$$

则

$$A^HA=R_1^HR_1=R_2^HR_2$$

由 $A\in C_r^{m\times r}$ 知 A^HA 正定。由§3.4知，正定矩阵 A^HA 的这种楚来斯基分解是唯一的，故 $R_1=R_2$。又由于 R_1 与 R_2 可逆，故由(1)式得 $Q_1=Q_2$，于是这种 QR 分解的唯一性得证。

推论 2 若 $A\in C_r^{r\times n}$，则 A 可唯一地分解为

$$A=LQ$$

其中 $Q\in C_r^{r\times n}$且 $QQ^H=I_r$，$L\in C_r^{r\times r}$为具有正对角元的下三角阵。

证明可仿推论1的证明。

由以上结果得知，矩阵 A 的 QR 分解中的矩阵 Q，显然是次酉矩阵。

例 用 QR 方法解方程组

$$Ax=b$$

其中

$$A=\begin{bmatrix}1&1&2\\1&2&1\\1&1&3\\2&3&3\end{bmatrix}=(\alpha_1\alpha_2\alpha_3)$$

$\alpha_1=(1,1,1,2)^T$，$\alpha_2=(1,2,1,3)^T$，$\alpha_3=(2,1,3,3)^T$

$x=(x_1,x_2,x_3)^T$, $b=(1,0,2,1)^T$

解 将 α_1、α_2、α_3 标准正交化为

$$\beta_1=\left(\frac{1}{7},\frac{1}{7},\frac{1}{7},\frac{2}{7}\right)^T,\ \beta_2=\left(-\frac{3}{\sqrt{35}},\frac{4}{\sqrt{35}},-\frac{3}{\sqrt{35}},\frac{1}{\sqrt{35}}\right)^T,\ \beta_3=\left(-\frac{2}{\sqrt{15}},\frac{1}{\sqrt{15}},\frac{3}{\sqrt{15}},-\frac{1}{\sqrt{15}}\right)^T$$

于是令

$$Q=\begin{bmatrix}\frac{1}{7} & -\frac{3}{\sqrt{35}} & -\frac{2}{\sqrt{15}}\\ \frac{1}{7} & \frac{4}{\sqrt{35}} & \frac{1}{\sqrt{15}}\\ \frac{1}{7} & -\frac{3}{\sqrt{35}} & \frac{3}{\sqrt{15}}\\ \frac{2}{7} & \frac{1}{\sqrt{35}} & -\frac{1}{\sqrt{15}}\end{bmatrix}$$

得

$$R=Q^HA=\begin{bmatrix}1 & \frac{10}{7} & \frac{12}{7}\\ 0 & \frac{5}{\sqrt{35}} & -\frac{8}{\sqrt{35}}\\ 0 & 0 & \frac{3}{\sqrt{15}}\end{bmatrix}$$

$$R^{-1}=\begin{bmatrix}1 & -\frac{2\sqrt{35}}{7} & -\frac{4\sqrt{15}}{3}\\ 0 & \frac{\sqrt{35}}{5} & \frac{8\sqrt{15}}{15}\\ 0 & 0 & \frac{\sqrt{15}}{3}\end{bmatrix}$$

于是

$$x = R^{-1}Q^H b = (-1, 0, 1)^T$$

§3.7 矩阵的奇值分解

矩阵的约当标准形的重要性已为人们所公认，但它有两点局限，一是它只是方阵的一种分解；二是虽然约当标准形是一特殊的上三角阵，但它仍不能象对角阵那样方便。人们的研究突破了这两点，获得了新的矩阵的奇值分解，它在现代矩阵理论中的重要性是不言而喻的，近年来，古典控制中的频率法，正是由于有了矩阵奇值分解的帮助而得到了新的发展。这里只给出分解定理。

定理 1 设 $A \in C^{m \times n}$，则

a) $A^H A$、AA^H 的特征值均为非负实数；

b) $A^H A$ 与 AA^H 的非零特征值相同。

证 a) 设 $\theta \neq x \in C^n$ 为 $A^H A$ 的特征值 λ 所对应的特征向量，$A^H A$ 为埃尔米特阵，故 λ 为实数，且

$$0 \leqslant (Ax, Ax) = (x, A^H Ax) = (x, \lambda x)$$
$$= \lambda(x, x)$$

由于 $x \neq \theta$，故 $(x, x) > 0$，所以 $\lambda \geqslant 0$；

相仿可知 AA^H 的特征值也是非负实数。

b) 设 $A^H A$ 的特征值依大小顺序编号为

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$$

而 AA^H 的特征值也依大小顺序编号为

$$\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_t > \mu_{t+1} = \mu_{t+2} = \cdots = \mu_m = 0$$

设 $\theta \neq x_i \in C^n (i \in \underline{r})$ 为 $A^H A$ 的非零特征值 $\lambda_i (i \in \underline{r})$ 所对应的特征向量，则由

$$A^H A x_i = \lambda_i x_i \qquad i \in \underline{r}$$

有

$$(AA^HA)x_i=\lambda_iAx_i \qquad i\in \underline{r}$$

且 $Ax_i\neq\theta$，于是 λ_i 也是 AA^H 的非零特征值。同理可证 AA^H 的非零特征值也是 A^HA 的非零特征值，如果还能证明 A^HA 与 AA^H 的非零特征值的代数重复度亦相同，则 A^HA 与 AA^H 的非零特征值就全同了。为此，设 y_1、y_2、…、y_p 为 A^HA 对应于特征值 $\lambda\neq0$的线性无关的特征向量，由于 A^HA 为单纯阵，故 p 即 λ 的代数重复度。显然，$Ay_i(i\in \underline{p})$是 AA^H 对应于 $\lambda\neq0$的特征向量。为证这些特征向量线性无关，令

$$k_1Ay_1+k_2Ay_2+\cdots+k_pAy_p=\theta$$

即

$$A(y_1y_2\cdots y_p)k=\theta$$

其中 $k=(k_1,k_2,\cdots,k_p)^T$。于是

$$A^HA(y_1,y_2,\cdots,y_p)k=\theta$$

即

$$\lambda(y_1,y_2,\cdots,y_p)k=\theta$$

已知 $\lambda\neq0$，故

$$(y_1,y_2,\cdots,y_p)k=\theta$$

已知 y_1、y_2、…、y_p 线性无关，故 $k=\theta$，即 Ay_1、Ay_2、…、Ay_p 线性无关，因而 λ 也是 AA^H 的 p 重非零特征值。证毕。

定义1 设 $A\in C_r^{m\times n}$，A^HA 的特征值为

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_r>\lambda_{r+1}=\lambda_{r+2}=\cdots=\lambda_n=0$$

则称 $\sigma_i=\sqrt{\lambda_i}\,(i\in \underline{r})$为矩阵 A 的正奇值。

由此定义可知，A 与 A^H 有相同的正奇值。

例1 设

$$A=\begin{bmatrix}1 & 2\\0 & 0\\0 & 0\end{bmatrix}$$

求 A 的正奇值。

解 由于

$$AA^H=\begin{bmatrix}1 & 2\\0 & 0\\0 & 0\end{bmatrix}\begin{pmatrix}1 & 0 & 0\\2 & 0 & 0\end{pmatrix}=\begin{bmatrix}5 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}$$

显然，AA^H 的正特征值为5，故 A 的正奇值为$\sqrt{5}$。

定义2 设 A，$B\in C^{m\times n}$，若存在 $S\in U^{m\times m}$，$T\in U^{n\times n}$，使得

$$A=SBT$$

则称 A 与 B **酉等价**。

定理2 若 A，$B\in C^{m\times n}$酉等价，则 A 与 B 有相同的正奇值。

证 由于 A 与 B 酉等价，故 AA^H 与 BB^H 酉相似，于是 AA^H、BB^H 有相同的特征值，因此 A 与 B 有相同的正奇值。

定理3 设 $A\in C_r^{m\times n}$，则存在 $U\in U^{m\times m}$，$V\in U^{n\times n}$，使得

$$A=U\begin{pmatrix}\Delta & O\\O & O\end{pmatrix}V^H \tag{1}$$

其中 $\Delta=\text{diag}(\delta_1, \delta_2, \cdots, \delta_r)$，$\delta_1$，$\delta_2$，$\cdots$，$\delta_r$ 是一组复数，且

$$|\delta_i|=\sigma_i \qquad i\in \underline{r}$$

$\sigma_i(i\in \underline{r})$是 A 的正奇值。

A 的式(1)形式的分解称为 A 的**奇值分解**，它表明 A 与一长方对角阵酉等价。

证 由于 AA^H 是正规阵，故存在$U\in U^{m\times m}$，使得

$$U^HAA^HU=\begin{pmatrix}\Delta\Delta^H & O\\O & O\end{pmatrix}$$

若记 $U=(U_1U_2)$，$U_1\in U^{m\times r}$，$U_2\in U^{m\times(m-r)}$，由于 $AA^HU_2=O$，因而有

$$A^H U_2 = O_{m\times(m-r)},\quad U_2^H A = O_{(m-r)\times m}$$

令 $V_1 = A^H U_1 \Delta^{-H}$，则有

$$V_1^H V_1 = \Delta^{-1} U_1^H A A^H U_1 \Delta^{-H} = I_r$$

由此可见 $V_1 \in U^{n\times r}$，令 $V_2 \in U^{n\times(n-r)}$，使 $V=(V_1 V_2)\in U^{n\times n}$，则有

$$V_1^H V_2 = \Delta^{-1} U_1^H A V_2 = O_{r\times(n-r)}$$

从而有

$$U_1^H A V_2 = O_{r\times(n-r)}$$

故有

$$U^H A V = \begin{pmatrix} U_1^H \\ U_2^H \end{pmatrix} A (V_1 V_2) = \begin{pmatrix} \Delta & O \\ O & O \end{pmatrix}$$

即(1)式成立。

例2 求例1中矩阵 A 的奇值分解。

解 由例1已知 A 的正奇值 $\sigma_1=\sqrt{5}$，则 $\Delta=(\sqrt{5})$，且由

$$AA^H = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

知 AA^H 的特征值 $\lambda_1=5$，$\lambda_2=\lambda_3=0$，对应的特征向量可分别取为

$$x_1 = e_1 = (1,0,0)^T,\quad x_2 = e_2 = (0,1,0)^T$$

$$x_3 = e_3 = (0,0,1)^T$$

则 $U=(e_1,e_2,e_3)=I_3$，其中 $U_1=(e_1)$，$U_2=(e_2,\ e_3)$，$V_1=$

$$A^H U_1 \Delta^{-H} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \frac{1}{\sqrt{5}} = \left(\frac{1}{\sqrt{5}},\ \frac{2}{\sqrt{5}} \right)^T$$，故可取

$$V_2=\begin{bmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$$

于是

$$A=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

定理4 （极分解）设 $A\in C^{n\times n}$，则可分解为

$$A=GU=UH$$

其中 G、H 为半正定 Hermite 矩阵，U 为酉阵。

证 由定理3知

$$A=U_1\begin{pmatrix} \Delta & O \\ O & O \end{pmatrix}V_1^H=U_1\begin{pmatrix} \Delta & O \\ O & O \end{pmatrix}U_1^HU_1V_1^H$$

$$=U_1V_1^HV_1\begin{pmatrix} \Delta & O \\ O & O \end{pmatrix}V_1^H$$

令

$$G=U_1\begin{pmatrix} \Delta & O \\ O & O \end{pmatrix}U_1^H,\quad H=V_1\begin{pmatrix} \Delta & O \\ O & O \end{pmatrix}V_1^H,$$

$$U=U_1V_1^H$$

即得，显然

$$UV^H=U_1V_1^HV_1U_1^H=U_1U_1^H=I$$

由于篇幅所限，还有一些重要的矩阵分解未能介绍。例如，程指军同志证明了任意域上的方阵，都可分解为两个对称矩阵的乘积等①。

---

① 《数学学报》第28卷第4期，1985年7月。

习 题 三

一、用克劳特法求解方程组：

$$\begin{cases} x_1+x_2+2x_3+3x_4=1 \\ 3x_1-x_2-x_3-2x_4=-4 \\ 2x_1+3x_2-x_3-x_4=-6 \\ x_1+2x_2+3x_3-x_4=-4 \end{cases}$$

二、求下列矩阵的约当标准形及相似变换矩阵：

1. $$A=\begin{bmatrix} 3 & 7 & -3 \\ -2 & -5 & 2 \\ -4 & -10 & 3 \end{bmatrix}$$

2. $$A=\begin{bmatrix} 4 & 5 & -2 \\ -2 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix}$$

三、用求约当标准形解微分方程组：

$$\begin{cases} \dfrac{dx_1}{dt}=2x_1+x_2+x_3 \\ \dfrac{dx_2}{dt}=x_1+2x_2+x_3 \\ \dfrac{dx_3}{dt}=2x_1+2x_2 \end{cases}$$

四、试证：若 $A\in C^{n\times n}$的相异特征值为 $\lambda_i(i\in \underline{\sigma})$，$m_i$、$\alpha_i$ 分别为 λ_i的代数重复度与几何重复度，则

1. $\det A = \prod_{i=1}^{\sigma}\lambda_i^{m_i}$

2. $\operatorname{tr}A = \sum_{i=1}^{\sigma}m_i\lambda_i$

五、求单纯矩阵

$$A=\begin{bmatrix}-29 & 6 & 18\\ -20 & 5 & 12\\ -40 & 8 & 25\end{bmatrix}$$

的谱分解。

六、当 $\lambda_i(i\in\underline{n})$ 为正规阵 $A\in C^{n\times n}$ 的特征值时，证 $|\lambda_i|^2(i\in\underline{n})$ 是 A^HA 与 AA^H 的特征值。

七、设 $A,B\in R^{n\times n}$，$A=A^T$，$B=-B^T$，$C=A+B$，求证以下三个条件等价：

a) C 为实正规阵；

b) $AB=BA$

c) $AB=-(AB)^T$

八、设 $A=A^H\in C^{n\times n}$ 正定，则存在正数 α、β，使得

$$\alpha x^Hx\leqslant x^HAx\leqslant\beta x^Hx \qquad \forall\ x\in C^n$$

九、设 $A,B\in C^{n\times n}$ 均为埃尔米特矩阵，且 $A\geqslant B\geqslant O$，试证

$$\det A\geqslant\det B\geqslant 0$$

十、设 $A,B\in C^{n\times n}$ 均为正定埃尔米特矩阵，则 AB 为正定埃尔米特矩阵的充要条件为 A 与 B 可换。命题对半正定亦成立。

十一、设 $A,B\in C^{n\times n}$ 均为埃尔米特矩阵，且 B 正定，试证 $\lambda(AB^{-1})$ 是实数，且

$$\max_x\frac{x^HAx}{x^HBx}=\max_i\lambda_i(AB^{-1})$$

$$\min_x\frac{x^HAx}{x^HBx}=\min_i\lambda_i(AB^{-1})$$

十二、求可逆矩阵 T，使

$$T^{-1}\begin{pmatrix}2 & 1\\ -1 & 0\end{pmatrix}T=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$$

并求

$$\begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}^k$$

十三、设 $A=A^T\in R^{n\times n}$，若有 x_1，$x_2\in R^n$，使得

$$x_1^TAx_1>0，x_2^TAx_2<0$$

则存在 $\theta\neq x_0\in R^n$，使

$$x_0^TAx_0=0$$

十四、若 $A=A^H$，$B=B^H\in C^{n\times n}$正定，则

a）$A+B$ 正定；

b）$AB=(AB)^H$ 时，AB 正定。

十五、求

$$A=\begin{bmatrix} 2 & 1 & -2 & 3 \\ 0 & 4 & 1 & 1 \\ 2 & 5 & -2 & 4 \end{bmatrix}$$

的最大秩分解，$A=BC$，并计算

$$C^T(CC^T)^{-1}(B^TB)^{-1}B^T$$

十六、求十五题所给矩阵 A 的 QR 分解。

十七、求十五题所给矩阵 A 的奇值分解。

十八、试证：对任意的 A，$B\in C^{n\times n}$，AB 与 BA 有相同的特征值。

十九、设 $A\in C^{n\times n}$，且 $A^k=I_n$（k 为正整数）。问 A 可对角化吗？

二十、试证 $A\in C^{n\times n}$正规的充要条件为 A 可分解为 $A=SQ=QS$，其中 Q 为酉阵，S 为半正定 Hermite 矩阵。

第四章　向量与矩阵的范数

在线性代数计算及建立矩阵分析的基本概念等方面，向量与矩阵的范数起着奠基的作用，因此是十分重要的基本概念。本章将介绍向量与矩阵的范数及其性质，以及范数、测度在特征值估计中的应用。当然，从这里我们将获得具体的线性赋范空间模型。

§4.1　向量的范数

向量范数是用来刻画向量大小的一种度量，正如过去在内积空间中，用内积来定义向量的长度一样，线性空间中两个向量之间的“距离”，是与向量的长度有关的，但“距离”的含义是多种多样的。

定义1　设映射$\|\cdot\|:V(C)\to R$满足

a)正定条件：$\|x\|>0 \quad \forall\ \theta\neq x\in V(C)$

b)齐次条件：$\|\alpha x\|=|\alpha|\|x\|$

$\quad\forall\ \alpha\in C,\forall\ x\in V(C)$

c)三角不等式：$\|x+y\|\leqslant\|x\|+\|y\|$

$\qquad x,y\in V(C)$

则称映射$\|\cdot\|$为定义在线性空间$V(C)$上的一个**向量范数**。

例1　设$x=(x_1,x_2,\cdots,x_n)^T\in C^n$，则

$$\|x\|_1\triangleq\sum_{i=1}^{n}|x_i|$$

$$\|x\|_p \triangleq \sum_{i=1}^{n}(|x_i|^p)^{\frac{1}{p}} \qquad 1<p<+\infty$$

$$\|x\|_\infty \triangleq \max_{1\leqslant i\leqslant n}|x_i|$$

都是 C^n 上的向量范数。

证 对于 $\|x\|_1$, $\|x\|_\infty$ 满足向量范数定义是显然的，今只证 $\|x\|_p$ $1<p<+\infty$ 满足向量范数的定义：

a) 显然；

b) 对任意的 $\alpha\in C$, $x=(x_1,x_2,\cdots,x_n)^T\in C^n$，有

$$\|\alpha x\|_p = (\sum_{i=1}^{n}|\alpha x_i|^p)^{\frac{1}{p}} = (\sum_{i=1}^{n}|\alpha|^p|x_i|^p)^{\frac{1}{p}}$$

$$= |\alpha|(\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}} = |\alpha|\|x\|_p;$$

c) 引用汉尔登(Hölder)不等式

$$\sum_{i=1}^{n}|a_i||b_i| \leqslant (\sum_{i=1}^{n}|a_i|^p)^{\frac{1}{p}}(\sum_{i=1}^{n}|b_i|^q)^{\frac{1}{q}}$$

其中 p、q 为正实数，且 $\frac{1}{p}+\frac{1}{q}=1$，有

$$\sum_{i=1}^{n}(|x_i|+|y_i|)^p = \sum_{i=1}^{n}|x_i|(|x_i|+|y_i|)^{p-1}$$

$$+ \sum_{i=1}^{n}|y_i|(|x_i|+|y_i|)^{p-1}$$

$$\leqslant (\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}}\left[\sum_{i=1}^{n}(|x_i|+|y_i|)^{(p-1)q}\right]^{\frac{1}{q}}$$

$$+ (\sum_{i=1}^{n}|y_i|^p)^{\frac{1}{p}}\left[\sum_{i=1}^{n}(|x_i|+|y_i|)^{(p-1)q}\right]^{\frac{1}{q}}$$

$$= \left[(\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}} + (\sum_{i=1}^{n}|y_i|^p)^{\frac{1}{p}}\right]$$

$$\times\left[\sum_{i=1}^{n}(|x_i|+|y_i|)^{(p-1)q}\right]^{\frac{1}{q}}$$

由于$(p-1)q=p$，故上面的不等式即为

$$\left[\sum_{i=1}^{n}(|x_i|+|y_i|)^p\right]^{\frac{1}{p}}\leqslant\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}$$

亦即

$$\|x+y\|_p\leqslant\|x\|_p+\|y\|_p$$

其中$y=(y_1,y_2,\cdots,y_n)^T\in C^n$，故$\|x\|_p$是向量范数。

注意：当$p=1$时，$\|x\|_p$即$\|x\|_1$；当$p=2$时，$\|x\|_2$正是酉(欧氏)空间中向量的长度。又易知

$$(\|x\|_\infty)^p=(\max_i|x_i|)^p\leqslant\sum_{i=1}^{n}|x_i|^p$$
$$\leqslant n\max_i|x_i|^p=n\|x\|_\infty^p$$

从而得

$$\|x\|_\infty\leqslant\|x\|_p\leqslant n^{\frac{1}{p}}\|x\|_\infty\qquad p\geqslant1$$

于是有

$$\lim_{p\to+\infty}\|x\|_p=\|x\|_\infty$$

这说明例1所给出的三种向量范数$\|x\|_1$、$\|x\|_p(1<p<+\infty)$、$\|x\|_\infty$可以统一成一种向量范数

$$\|x\|_p\qquad 1\leqslant p\leqslant+\infty$$

并且例1也表明，即使在同一线性空间中，可根据不同的要求来定义向量的范数，即向量范数不唯一。

利用C^n与$V_n(C)$的同构关系，设$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$为$V_n(C)$的标准正交基，$\alpha=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)x\in V_n(C)$，$x=(x_1,x_2,\cdots,x_n)^T\in C^n$，则可以定义$\alpha\in V_n(C)$的$p$范数为

$$\|\alpha\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}\qquad 1\leqslant p\leqslant+\infty$$

例2 设 $f(t)$ 是 $[t_1,t_2]$ 上的任一连续函数，则

$$\|f(t)\|_1 \triangleq \int_{t_1}^{t_2} |f(t)| \mathrm{d}t$$

$$\|f(t)\|_p \triangleq \left(\int_{t_1}^{t_2} |f(t)|^p \mathrm{d}t\right)^{\frac{1}{p}} \qquad 1<p<+\infty$$

$$\|f(t)\|_\infty \triangleq \max_{t\in[t_1,t_2]} |f(t)|$$

都是 $[t_1,t_2]$ 上的连续函数所构成的线性空间上的范数。

$\|f(t)\|_1$ 与 $\|f(t)\|_\infty$ 是范数，显然。要证 $\|f(t)\|_p$ $(1<p<+\infty)$ 是范数，则要用到积分形式的汗尔登不等式

$$\int_{t_1}^{t_2} |f(t)g(t)| \mathrm{d}t \leqslant \left(\int_{t_1}^{t_2} |f(t)|^p \mathrm{d}t\right)^{\frac{1}{p}} \times \left(\int_{t_1}^{t_2} |g(t)|^q \mathrm{d}t\right)^{\frac{1}{q}}$$

其中 p、q 为正实数，$\frac{1}{p}+\frac{1}{q}=1$。

证明留给读者自己练习。

例3 设 $A=(a_{ij})\in R^{m\times n}$（或 $C^{m\times n}$），令

$$\|A\|_{v_1} \triangleq \sum_{j=1}^{n}\sum_{i=1}^{m} |a_{ij}|$$

$$\|A\|_{v_p} \triangleq \left(\sum_{j=1}^{n}\sum_{i=1}^{m} |a_{ij}|^p\right)^{\frac{1}{p}} \quad 1<p<+\infty$$

$$\|A\|_{v_\infty} \triangleq \max_{i,j} |a_{ij}| \quad i\in \underline{m}, j\in \underline{n}$$

易证它们都是线性空间 $R^{m\times n}$（或 $C^{m\times n}$）上的向量范数。特别当 $p=2$ 时，称

$$\|A\|_F \triangleq \left(\sum_{j=1}^{n}\sum_{i=1}^{m} |a_{ij}|^2\right)^{\frac{1}{2}}$$

为 A 的弗罗贝纽斯(Frobenius)范数，它是最常用的范数之一。

例4 设 $x\in R^1$，若令 $\|x\|=x^2$，虽然它满足定义1中 a)，但不满足定义1中 b)，故它不是 R^1 中的范数。

例5 在 R^n(或 C^n)中，若 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^T$，令

$$\|\boldsymbol{x}\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}\qquad 0<p<1$$

由于它不满足定义1中 c)，故它不是 R^n(或 C^n)上的向量范数，例如，在 R^2 中，取 $\boldsymbol{x}=(1,0)^T$，$\boldsymbol{y}=(0,1)^T$，则

$$\|\boldsymbol{x}+\boldsymbol{y}\|_{\frac{1}{2}}=4,\quad \|\boldsymbol{x}\|_{\frac{1}{2}}=1,\quad \|\boldsymbol{y}\|_{\frac{1}{2}}=1$$

故 $\|\;\|_{\frac{1}{2}}$ 不是 R^2 上的向量范数。

定义了向量范数 $\|\cdot\|$ 的线性空间 $V_n(F)$，称为线性赋范空间，记为 $V_n(F,\|\quad\|)$。

向量范数具有以下性质：

定理1 设 $\|\boldsymbol{x}\|$ 是 $V_n(F)$ 上的向量范数，则

a) $\|\theta\|=0$

b) $\|\boldsymbol{x}-\boldsymbol{y}\|\geqslant|\|\boldsymbol{x}\|-\|\boldsymbol{y}\||\quad \forall\ \boldsymbol{x},\boldsymbol{y}\in V_n(F)$ (1)

c) 设 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 为 $V_n(F)$ 的标准正交基，$\boldsymbol{x}=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\tilde{x}$，$\tilde{x}=(x_1,x_2,\cdots,x_n)^T\in F^n$，则 $\|x\|$ 在闭球

$$S=\{\boldsymbol{x}\mid(\tilde{x},\tilde{x})^{\frac{1}{2}}\leqslant 1\}$$

上有界；

d) $\|\boldsymbol{x}\|$ 是关于 $\|\boldsymbol{x}\|_2$ 的连续函数。

证

a) $\|\theta\|=\|0\boldsymbol{x}\|=0\|\boldsymbol{x}\|=0$

b) 对 $\boldsymbol{x}=\boldsymbol{y}+(\boldsymbol{x}-\boldsymbol{y})$ 使用定义1中的"三角不等式"，有

$$\|\boldsymbol{x}-\boldsymbol{y}\|\geqslant\|\boldsymbol{x}\|-\|\boldsymbol{y}\|\tag{2}$$

同理有

$$\|\boldsymbol{y}-\boldsymbol{x}\|\geqslant\|\boldsymbol{y}\|-\|\boldsymbol{x}\|\tag{3}$$

又

$$\|x-y\| = \|-(y-x)\| = \|y-x\|$$

综合(2)与(3)，即知(1)成立；

c) 当 $x=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\tilde{x}\in S$ 时，则

$$(\tilde{x},\tilde{x}) = \sum_{i=1}^{n}|x_i|^2 \leqslant 1$$

所以 $|x_i|\leqslant 1(i\in \underline{n})$，于是，若记 $\sum\limits_{i=1}^{n}\|\varepsilon_i\|=M>0$，则

$$\|x\| = \left\|\sum_{i=1}^{n}x_i\varepsilon_i\right\| \leqslant \sum_{i=1}^{n}|x_i|\,\|\varepsilon_i\|$$

$$\leqslant \sum_{i=1}^{n}\|\varepsilon_i\| = M$$

即 $\|x\|$ 在 S 上有界；

d) 设 $\Delta\tilde{x}=(\Delta x_1,\Delta x_2,\cdots,\Delta x_n)^T\in F^n$，且

$$\Delta x = (\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\Delta\tilde{x} \in V_n(F)$$

则由 b)有

$$\big|\|x+\Delta x\| - \|x\|\big| \leqslant \|\Delta x\| = \|\Delta x\|_2\left\|\frac{\Delta x}{\|\Delta x\|_2}\right\|$$

由于 $\dfrac{\Delta x}{\|\Delta x\|_2}\in S$，故由 c)知 $\left\|\dfrac{\Delta x}{\|\Delta x\|_2}\right\|$ 有界，又

$$\lim_{\Delta x\to\theta}\|\Delta x\|_2 = 0$$

即

$$\lim_{\Delta x\to\theta}\big|\|x+\Delta x\| - \|x\|\big| = 0$$

故 d)得证。

定义2 设在 $V_n(F)$ 上定义了 $\|\cdot\|_a$，$\|\cdot\|_b$ 两种向量范数，若存在常数 $c_1>0$，$c_2>0$，使得

$$\|x\|_a \leqslant c_1\|x\|_b,\ \|x\|_b \leqslant c_2\|x\|_a\ \forall\ x \in V_n(F) \tag{4}$$

则称 $V_n(F)$ 的两个向量范数 $\|\cdot\|_a$ 与 $\|\cdot\|_b$ 等价。

定理2 $V_n(F)$上的任意两个向量范数均等价。

证 设$\|\cdot\|_a$、$\|\cdot\|_b$为$V_n(F)$上的任意两个向量范数，由定理1的d)知，对任意的$\theta \neq x \in V_n(F)$，映射

$$\varphi(x) = \frac{\|x\|_a}{\|x\|_b}$$

是$S=\{x \mid \tilde{x}^H\tilde{x}=1\}$上关于$\|x\|_2$的连续函数。由于$S$是有界闭集，故$\varphi(x)$在$S$上有界，即存在常数$k_1>0$，使得

$$\varphi(x) \leqslant k_1$$

即

$$\|x\|_a \leqslant k_1\|x\|_b$$

同理可证，存在常数$k_2>0$，使得

$$\|x\|_b \leqslant k_2\|x\|_a$$

因此$V_n(F)$上的任意两个向量范数等价。

注意：向量范数等价定义中的不等式((4)式)，也可改写为存在正常数$0<m<M$，使

$$m\|x\|_a \leqslant \|x\|_b \leqslant M\|x\|_a$$

向量范数的“等价”，是指有关按$\|\cdot\|_a$收敛的性质，按$\|\cdot\|_b$也相应成立。

§4.2 矩阵的范数

在§4.1例3中，给出的mn维线性空间$F^{m\times n}$($F=R$或$F=C$)上的向量范数$\|A\|_{v_1}$、$\|A\|_{v_p}$、$\|A\|_{v_\infty}$($A\in F^{m\times n}$这时A是作为向量看待的)，在线性空间中只需考虑加法运算与数乘运算。但若把A看作矩阵集合$F^{n\times n}$中的矩阵，还必须考虑矩阵的乘法运算，或者从矩阵分解的角度考虑，一个方阵的范数与其分解的各因子的范数之间，应有一种确定的关系。例如，为了用矩阵A、B的范数$\|A\|$、$\|B\|$来控制AB的范数$\|AB\|$，要

求有

$$\|AB\| \leqslant \|A\|\|B\| \qquad \forall A,B \in F^{n\times n}$$

对于$\|\cdot\|_{v_\infty}$就不一定满足上述不等式。例如取

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

显然 $AB=\begin{pmatrix} 2 & 0 \\ 1 & 0 \end{pmatrix}$，$\|A\|_{v_\infty}=1$，$\|B\|_{v_\infty}=1$，$\|AB\|_{v_\infty}=2$上述不等式就不能满足。

定义 设 $A,B\in F^{n\times n}$，若映射$\|\cdot\|:F^{n\times n}\to R$，且满足

$a)\|A\|>0 \qquad \forall O\neq A\in F^{n\times n}$

$b)\|\alpha A\|=|\alpha|\|A\| \qquad \forall \alpha\in F,\forall A\in F^{n\times n}$

$c)\|A+B\|\leqslant\|A\|+\|B\| \qquad \forall A,B\in F^{n\times n}$

$d)\|AB\|\leqslant\|A\|\|B\| \qquad \forall A,B\in F^{n\times n}$

则称映射$\|\cdot\|$为 $F^{n\times n}$上的矩阵范数。

例 设 $A=(a_{ij})\in F^{n\times n}$，若

$$\|A\|_{m_1}\triangleq\sum_{j=1}^{n}\sum_{i=1}^{n}|a_{ij}|$$

$$\|A\|_{m_2}\triangleq\Big(\sum_{j=1}^{n}\sum_{i=1}^{n}|a_{ij}|^2\Big)^{\frac{1}{2}}$$

$$\|A\|_{m_\infty}\triangleq n\max_{ij}|a_{ij}|$$

则它们都是矩阵范数。

证 由于矩阵范数定义的前三条与向量范数定义的三条相似，故它们满足矩阵范数定义的前三条是显然的，现只证它们满足矩阵范数定义的第四条。

$$\|AB\|_{m_1}=\sum_{j=1}^{n}\sum_{i=1}^{n}\Big|\sum_{k=1}^{n}a_{ik}b_{kj}\Big|\leqslant\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}||b_{kj}|$$

$$\leqslant\Big(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|\Big)\Big(\sum_{k=1}^{n}\sum_{j=1}^{n}|b_{kj}|\Big)=\|A\|_{m_1}\|B\|_{m_1}$$

$$\|AB\|_{m_2}=\left(\sum_{j=1}^{n}\sum_{i=1}^{n}\left|\sum_{k=1}^{n}a_{ik}b_{kj}\right|^2\right)^{\frac{1}{2}}$$

$$\leqslant\left[\sum_{j=1}^{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}||b_{kj}|\right)^2\right]^{\frac{1}{2}}$$

$$\leqslant\left[\sum_{j=1}^{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}|^2\right)\left(\sum_{k=1}^{n}|b_{kj}|^2\right)\right]^{\frac{1}{2}}$$

$$=\left(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|^2\right)^{\frac{1}{2}}\left(\sum_{k=1}^{n}\sum_{j=1}^{n}|b_{kj}|^2\right)^{\frac{1}{2}}$$

$$=\|A\|_{m_2}\|B\|_{m_2}$$

$$\|AB\|_{m_\infty}=n\max_{i\,j}\left|\sum_{k=1}^{n}a_{ik}b_{kj}\right|$$

$$\leqslant n\max_{i\,j}\sum_{k=1}^{n}|a_{ik}||b_{kj}|$$

$$\leqslant n\max_{i\,j}n\max_{k}|a_{ik}||b_{kj}|$$

$$\leqslant n\max_{i\,j}|a_{ik}|n\max_{k,j}|b_{kj}|=\|A\|_{m_\infty}\|B\|_{m_\infty}$$

与向量范数相似，矩阵范数也具有相应性质，由下述定理给出。

定理1 设 $A,B\in F^{n\times n}$，则

a） $\|O_{n\times n}\|=0$

b） $\|A-B\|\geqslant|\|A\|-\|B\||$

c） $\|A\|$在 $S=\{A\mid\|A\|_{m_2}\leqslant1\}$上有界；

d） $\|A\|$关于$\|A\|_{m_2}$是连续函数；

e） $F^{n\times n}$上的任意两个矩阵范数等价。

这个定理的证明与上节定理1的证明类似。

在矩阵范数中，$\|A\|_{m_2}$用途最多，它具有下述性质。

定理2 设 $A\in F^{n\times n}$，

a） 若 $A=(\alpha_1,\alpha_2,\cdots,\alpha_n)$，则

$$\|A\|_{m_2}^2=\sum_{i=1}^{n}\|\alpha_i\|_2^2$$

其中 $\|\alpha_i\|_2^2=\alpha_i^H\alpha_i$ 是 F^n 中的向量范数。

b) $\|A\|_{m_2}^2=\mathrm{tr}A^HA=\sum_{i=1}^{n}\lambda_i(A^HA)$

c) 对任意的 U、$V\in U^{n\times n}$，有

$$\|A\|_{m_2}=\|U^HAV\|_{m_2}=\|UAV^H\|_{m_2}$$

证 a)、b)显然；

c) $\|A\|_{m_2}=\mathrm{tr}A^HA=\mathrm{tr}(A^HUU^HA)$

$=\mathrm{tr}(V^HA^HUU^HAV)=\|U^HAV\|_{m_2}$

同理可证

$$\|A\|_{m_2}=\|UAV^H\|_{m_2}$$

§4.3 算子范数

在矩阵理论中，矩阵 $A\in F^{m\times n}$ 除了可作为线性空间 $F^{m\times n}$ 中的向量外，也可将它视为线性映射(或线性算子) $A:F^n\to F^m$，即

$$Ax\in F^m \qquad \forall\, x\in F^n$$

而矩阵范数与向量范数又是有差异的，若将向量视作矩阵，则应有

$$\|Ax\|\leqslant\|A\|\|x\|$$

这个不等式中既有向量也有矩阵，要区别矩阵与向量，它们各自取的范数仍能使这个不等式保持吗?若能保持就称它们相容。

定义 若 $\|\cdot\|$ 是 F^n 的向量范数，$\|\cdot\|_m$ 是 $F^{n\times n}$ 的矩阵范数，且

$$\|Ax\|\leqslant\|A\|_m\|x\| \qquad A\in F^{n\times n},x\in F^n \tag{1}$$

则称$\|\cdot\|_m$为与向量范数$\|\cdot\|$相容的矩阵范数。

例1 设$x\in F^n, A=(a_{ij})\in F^{n\times n}$，若

$$\|x\|_1=\sum_{i=1}^{n}|x_i| \qquad x=(x_1,x_2,\cdots,x_n)^T$$

则

$$\|A\|_{m_1}=\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|$$

可作为与$\|\cdot\|_1$相容的矩阵范数。

证 已知$\|A\|_{m_1}$为矩阵范数，故只须证它与$\|\cdot\|_1$满足不等式(1)即可。因为

$$\|Ax\|_1=\sum_{i=1}^{n}\left|\sum_{k=1}^{n}a_{ik}x_k\right|\leqslant\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}||x_k|$$

$$\leqslant\left(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ij}|\right)\sum_{j=1}^{n}|x_j|=\|A\|_{m_1}\|x\|_1$$

故$\|A\|_{m_1}$与$\|x\|_1$相容。

当向量范数和矩阵给定时，与该向量范数相容的矩阵范数总是存在的，这可由下述定理给出。

定理1 设$\|x\|$为F^n上的向量范数，$A\in F^{n\times n}$，则

$$\|A\|=\max_{x\neq\theta}\frac{\|Ax\|}{\|x\|}\left(=\max_{\|x\|=1}\|Ax\|\right) \tag{2}$$

是与向量范数$\|x\|$相容的矩阵范数。

证 此时不等式(1)是显然成立的，故只须证$\|A\|$确是矩阵范数即可。

a)设$A\neq O_{n\times n}$，则存在$\theta\neq x_0\in F^n$，使

$$Ax_0\neq\theta$$

于是

$$\|A\|\geqslant\frac{\|Ax_0\|}{\|x_0\|}>0$$

$$b)\ \|\alpha A\|=\max_{x\neq\theta}\frac{\|\alpha Ax\|}{\|x\|}=\max_{x\neq\theta}\frac{|\alpha|\,\|Ax\|}{\|x\|}$$

$$=|\alpha|\max_{x\neq\theta}\frac{\|Ax\|}{\|x\|}=|\alpha|\,\|A\|$$

$$c)\ \|A+B\|=\max_{x\neq\theta}\frac{\|(A+B)x\|}{\|x\|}$$

$$\leqslant\max_{x\neq\theta}\frac{\|Ax\|+\|Bx\|}{\|x\|}\leqslant\max_{x\neq\theta}\frac{\|Ax\|}{\|x\|}$$

$$+\max_{x\neq\theta}\frac{\|Bx\|}{\|x\|}=\|A\|+\|B\|$$

$$d)\ \|AB\|=\max_{x\neq\theta}\frac{\|ABx\|}{\|x\|}=\max_{\substack{x\neq\theta\\ Bx\neq\theta}}\frac{\|A(Bx)\|}{\|Bx\|}\ \frac{\|Bx\|}{\|x\|}$$

$$\leqslant\max_{Bx\neq\theta}\frac{\|A(Bx)\|}{\|Bx\|}\max_{x\neq\theta}\frac{\|Bx\|}{\|x\|}$$

$$=\|A\|\,\|B\|$$

证毕。

例2 与向量范数$\|x\|_1=\sum_{i=1}^{n}|x_i|$相容的矩阵范数也可取为

$$\|A\|_1=\max_{j}\sum_{i=1}^{n}|a_{ij}|$$

解 $\|Ax\|_1=\sum_{i=1}^{n}\left|\sum_{j=1}^{n}a_{ij}x_j\right|\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|\,|x_j|$

$$=\sum_{j=1}^{n}\left(\sum_{i=1}^{n}|a_{ij}|\right)|x_j|$$

$$\leqslant\max_{j}\sum_{i=1}^{n}|a_{ij}|\sum_{j=1}^{n}|x_j|=\|A\|_1\|x\|_1$$

故$\|A\|_1$与$\|x\|_1$相容。若令

$$\lambda=\sum_{i=1}^{n}|a_{il}|=\max_{j}\sum_{i=1}^{n}|a_{ij}|\qquad 1\leqslant l\leqslant n$$

记 $A=(a_1,a_2,\cdots,a_n)$，则 $\lambda=\|a_l\|_1$，又取

$$e_l=(\overbrace{0,\cdots,0,1}^{\text{第}l\text{个}},0,\cdots,0)^T \qquad 1\leqslant l\leqslant n$$

于是

$$\|Ae_l\|_1=\|a_l\|_1=\lambda\|e_l\|_1$$

故

$$\lambda=\max_{x\neq\theta}\frac{\|Ax\|_1}{\|x\|_1}=\max_j\sum_{i=1}^n|a_{ij}|=\|A\|_1$$

即$\|A\|_1$是对应于$\|x\|_1$的矩阵范数。

由此例及例1可知，$\|A\|_{m_1}$、$\|A\|_1$均与$\|x\|_1$相容。但$\|A\|_1$是与$\|x\|_1$相容的范数中值最小的一个，为区别于一般的相容范数，则称由等式(2)定义的相容范数为**算子范数**。于是$\|A\|_1$是对应于$\|x\|_1$的算子范数。

例3 证明对应于$\|x\|_2=\left(\sum\limits_{i=1}^n|x_i|^2\right)^{\frac{1}{2}}$的算子范数为：

$$\|A\|_2=\max_i\sqrt{\lambda_i(A^HA)}$$

证 当$x\neq\theta$时，因为

$$\frac{\|Ax\|_2}{\|x\|_2}=\frac{x^HA^HAx}{x^Hx}$$

由瑞利商的性质知

$$\max_{x\neq\theta}\frac{\|Ax\|_2}{\|x\|_2}=\max_j\lambda_j(A^HA)$$

故

$$\|A\|_2=\max_{x\neq\theta}\frac{\|Ax\|_2}{\|x\|_2}=\max_i\sqrt{\lambda_j(A^HA)}$$

例4 证明对应于$\|x\|_\infty=\max\limits_i|x_i|$的算子范数为

$$\|A\|_\infty=\max_i\sum_{j=1}^n|a_{ij}|$$

证 设 $\mu=\sum_{j=1}^{n}|a_{lj}|=\max_{i}\sum_{j=1}^{n}|a_{ij}|, 1\leqslant l\leqslant n$，则

$$\|Ax\|_{\infty}=\max_{i}\left|\sum_{k=1}^{n}a_{ik}x_{k}\right|\leqslant\max_{i}\sum_{k=1}^{n}|a_{ik}||x_{k}|$$

$$\leqslant\max_{i}\sum_{k=1}^{n}|a_{ik}|\max_{k}|x_{k}|=\mu\|x\|_{\infty}$$

即 $\|A\|_{\infty}$ 与 $\|x\|_{\infty}$ 相容。

其次，若记 $a_{lj}=|a_{lj}|e_{j}^{i\theta}$，其中 $\theta_{j}\in R, j\in\underline{n}, i=\sqrt{-1}$，则令 $z=(e^{-i\theta_{1}},e^{-i\theta_{2}},\cdots,e^{-i\theta_{n}})^{T}$，有 $\|z\|_{\infty}=1$，但

$$\|Az\|_{\infty}\geqslant\left|\sum_{j=1}^{n}a_{lj}e^{-i\theta_{j}}\right|=\sum_{j=1}^{n}|a_{lj}|=\mu=\mu\|z\|_{\infty}$$

综合以上结果可知，$\|A\|_{\infty}$ 为对应于 $\|x\|_{\infty}$ 的算子范数。

最常用的算子范数 $\|A\|_{2}$，具有以下性质。

定理2 设 $A\in F^{n\times n}$，则

a) $\|A\|_{2}=\max_{i\,j}\{|y^{H}Ax|\mid\|x\|_{2}=\|y\|_{2}=1, x,y\in F^{n}\}$

b) $\|A\|_{2}=\|\overline{A}\|_{2}=\|A^{T}\|_{2}=\|A^{H}\|_{2}$

c) $\|A^{H}A\|_{2}=\|A\|_{2}^{2}$

d) 对于任意的 $U,V\in U^{n\times n}$，有

$$\|A\|_{2}=\|UAV\|_{2}$$

证 a)由于

$$|y^{H}Ax|=(y,Ax)\leqslant\|y\|_{2}\|Ax\|_{2}$$

$$\leqslant\|A\|_{2}\|x\|_{2}\|y\|_{2}=\|A\|_{2}$$

故 $\|A\|_{2}$ 为 $\|Ax\|_{2}$ 的最大值，因此存在 $x_{0}\neq\theta$，$\|x_{0}\|_{2}=1$，使得

$$\|Ax_{0}\|_{2}=\|A\|_{2}>0$$

取

$$y_{0}=\frac{Ax_{0}}{\|Ax_{0}\|_{2}}$$

则

$$|y_0^H A x_0| = \left|\frac{(Ax_0)^H}{\|Ax_0\|_2} A x_0\right| = \|Ax_0\|_2 = \|A\|_2$$

所以得

$$\|A\|_2 = \max_{xy}\{\,|y^H A x|\;|\;\|x\|_2 = \|y\|_2 = 1\}$$

b)因为

$$|y^H A x| = |\overline{y^H A x}| = |y^T \overline{A}\,\overline{x}|$$

$$= |\overline{y}^H \overline{A}\,\overline{x}| \quad \forall\ x, y \in F^n$$

又$\|y\|_2 = \|\overline{y}\|_2$，$\|x\|_2 = \|\overline{x}\|_2$，$S = \{x \,|\, \|x\|_2 = 1\}$时，

$$\|A\|_2 = \max_{xy}\{\,|y^H A x|\;|\;\|x\|_2 = \|y\|_2 = 1\}$$

$$= \max_{\overline{x}\overline{y}}\{\,|\overline{y}^H \overline{A}\,\overline{x}|\;|\;\|\overline{x}\|_2 = \|\overline{y}\|_2 = 1\} = \|\overline{A}\|_2$$

又由

$$|y^H A x| = |(y^H A x)^H| = |x^H A^H y|$$

知

$$\|A\|_2 = \max_{xy}\{\,|y^H A x|\;|\;\|x\|_2 = \|y\|_2 = 1\}$$

$$= \max_{xy}\{\,|x^H A^H y|\;|\;\|x\|_2 = \|y\|_2 = 1\} = \|A^H\|_2$$

故有

$$\|A\|_2 = \|\overline{A}\|_2 = \|A^H\|_2 = \|\overline{A}^H\|_2 = \|A^T\|_2$$

c)由矩阵范数定义及b)，有

$$\|A^H A\|_2 \leqslant \|A^H\|_2 \|A\|_2 = \|A\|_2^2$$

又由a)有

$$\|A^H A\|_2 = \max_{xy}\{\,|y^H A^H A x|\;|\;\|x\|_2 = \|y\|_2 = 1\}$$

$$\geqslant \max_{x}\{\,|x^H A^H A x|\;|\;\|x\|_2 = 1\} = \|A\|_2^2$$

综合以上结果即得

$$\|A^H A\|_2 = \|A\|_2^2$$

d)由 b)与 c)可得

$$\|A\|_2^2 = \|A^H A\|_2 = \|A^H U^H U A\|_2 = \|UA\|_2^2$$
$$= \|A^H U^H\|_2^2 = \|V^H A^H U^H\|_2^2 = \|UAV\|_2^2$$

其中 U、$V \in U^{n\times n}$。故

$$\|A\|_2 = \|UAV\|_2$$

以上讨论了当给定了向量范数和矩阵时，就能确定与给定的向量范数相容的矩阵范数；反之，当给定矩阵范数时，能否找到与其相容的向量范数？下面定理作出肯定的回答。

定理3　设 $\|A\|_*$ 是矩阵范数，则存在向量范数 $\|x\|$，使得

$$\|Ax\| \leqslant \|A\|_* \|x\|$$

证　任给 $\theta \neq a \in F^n$，定义

$$\|x\| = \|xa^H\|_* \qquad \forall\, x \in F^n$$

由于 $a \neq \theta$，故只要 $x \neq \theta$，就有 $xa^H \neq O_{n\times n}$，从而

$$\|x\| > 0 \qquad x \neq \theta$$

其次，任取 $\alpha \in F$，则

$$\|\alpha x\| = \|\alpha x a^H\|_* = |\alpha|\,\|xa^H\|_* = |\alpha|\,\|x\|$$

如果再设 $y \in F^n$，则

$$\|x + y\| = \|(x + y)a^H\|_* = \|xa^H + ya^H\|_*$$
$$\leqslant \|xa^H\|_* + \|ya^H\|_* = \|x\| + \|y\|$$

最后，再证 $\|x\|$ 与 $\|A\|_*$ 相容。因为

$$\|Ax\| = \|Axa^H\|_* \leqslant \|A\|_* \|xa^H\|_*$$
$$= \|A\|_* \|x\|$$

证毕。

例5　设 $\|A\|_* = \|A\|_{m_2}$，求一与其相容的向量范数。

解　取 $a = e_1 = (1, 0, \cdots, 0)^T$，$x = (x_1, x_2, \cdots, x_n)^T \in C^n$，于是由定理3，有

$$\|x\| = \|xa^H\|_{m_2} = \left\|\begin{pmatrix} x_1 & 0 & 0 & \cdots & 0 \\ x_2 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_n & 0 & 0 & \cdots & 0 \end{pmatrix}\right\|_{m_2}$$

$$= \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}} = \|x\|_2$$

§4.4 矩阵的测度

与矩阵范数一样，在系统理论中，还有一个经常用到的量，称为矩阵的测度(measure)。由于其定义与范数密切相关，所以在讨论微分方程的解的稳定性、大范围平衡点的存在性、网络数值解误差的评价等时，利用测度比利用范数往往会得到更为明确的条件和评价。

定义 设 $A \in F^{n\times n}$，$\|A\|$是给定的算子范数，若

$$\mu(A) = \lim_{h\to 0+} \frac{\|I_n + hA\| - 1}{h} \tag{1}$$

存在，则称 $\mu(A)$为对应于$\|A\|$的 A 的测度。

由于$\|I_n\|=1$，因此，(1)式可视为$\|\cdot\|$在 I_n 点沿 A 方向的方向导数。其次，即使是同一矩阵 A，由于所给算子范数不同，测度 $\mu(A)$的值亦可能不同，即 $\mu(A)$是与一确定的$\|A\|$相对应的。

例1 算子范数$\|A\|_1$对应的矩阵测度

$$\mu_1(A) = \max_j\left[\mathrm{Re}\,a_{jj} + \sum_{\substack{i=1\\ i\neq j}}^{n} |a_{ij}|\right] \tag{2}$$

算子范数$\|A\|_\infty$对应的矩阵测度

$$\mu_\infty(A) = \max_i\left[\mathrm{Re}\,a_{ii} + \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|\right] \tag{3}$$

证 由定义

$$\mu_1(A)=\lim_{h\to 0+}\frac{\|I_n+hA\|_1-1}{h}$$

$$=\lim_{h\to 0+}\frac{\max_j\left[\sum_{i=1}^{n}|\delta_{ij}+ha_{ij}|\right]-1}{h}$$

$$=\lim_{h\to 0+}\frac{\max_j\left[\,|1+ha_{jj}|+h\sum_{\substack{i=1\\i\neq j}}^{n}|a_{ij}|\right]-1}{h}$$

$$=\lim_{h\to 0+}\frac{\max_j\left[\,|1+h\mathrm{Re}a_{jj}+o(h)+h\sum_{\substack{i=1\\i\neq j}}^{n}|a_{ij}|\right]-1}{h}$$

$$=\max_j\left[\mathrm{Re}a_{jj}+\sum_{\substack{i=1\\i\neq j}}^{n}|a_{ij}|\right]$$

同理可证(3)式成立。

例2 算子范数$\|A\|_2$对应的矩阵测度

$$\mu_2(A)=\max_i\lambda_i\left(\frac{A^H+A}{2}\right) \tag{4}$$

证 $\mu_2(A)=\lim\limits_{h\to 0+}\dfrac{\|I_n+hA\|_2-1}{h}$

$$=\lim_{h\to 0+}\frac{\max_i\sqrt{\lambda_i[(I_n+hA^H)(I_n+hA)]}-1}{h}$$

$$=\lim_{h\to 0+}\frac{\max_i\sqrt{\lambda_i\left[I_n+2h\dfrac{A^H+A}{2}+h^2A^HA\right]}-1}{h}$$

$$=\lim_{i}\frac{\max_i\left[1+h\lambda_i\left(\dfrac{A^H+A}{2}\right)+o(h)\right]-1}{h}$$

$$=\max_i\lambda_i\left(\frac{A^H+A}{2}\right)$$

例3 若 $A=A^H\in C^{n\times n}$，则

a) $\|A\|_2=\max\limits_i|\lambda_i(A)|$ (5)

b) $\mu_2(A)=\max\limits_i\lambda_i(A)$ (6)

c) $-\mu_2(-A)=\min\limits_i\lambda_i(A)$ (7)

证 a) $\|A\|_2=\max\limits_i\sqrt{\lambda_i(A^HA)}=\max\limits_i\sqrt{\lambda_i(A^2)}$

$=\max\limits_i|\lambda_i(A)|$

b) $\mu_2(A)=\max\limits_i\lambda_i\left(\dfrac{A^H+A}{2}\right)=\max\limits_i\lambda_i(A)$

c) $-\mu_2(-A)=-\max\limits_i\lambda_i(-A)=-\max\limits_i[-\lambda_i(A)]$

$=-[-\min\limits_i\lambda_i(A)]=\min\limits_i\lambda_i(A)$

矩阵的测度有下列性质：

定理 设 $A\in F^{n\times n}$，$\|A\|$为算子范数，则

a) $\mu(I_n)=1,\mu(-I_n)=-1,\mu(O_{n\times n})=0$

b) $\mu(\alpha A)=\alpha\mu(A),\qquad 0\leqslant\alpha\in F$

c) $\mu(A+\alpha I_n)=\mu(A)+\alpha\quad\forall\ \alpha\in F$

d) $\mu(A+B)\leqslant\mu(A)+\mu(B)\qquad\forall\ A,B\in F^{n\times n}$

e) $-\|A\|\leqslant-\mu(-A)\leqslant\mu(A)\leqslant\|A\|$ (8)

f) $-\mu(-A)\leqslant\operatorname{Re}\lambda_i(A)\leqslant\mu(A)\qquad i\in\underline{n}$

g) $\|Ax\|\geqslant\max\{-\mu(-A),-\mu(A)\}\|x\|$

$\forall\ x\in F^n$ (9)

证 a)、b)由矩阵测度的定义显然成立；

c) $\mu(A+\alpha I_n)=\lim\limits_{h\to0+}\dfrac{\|I_n+h(A+\alpha I_n)\|-1}{h}$

$$=\lim_{h\to0+}\frac{\left\|I_n+\dfrac{h}{1+\alpha h}A\right\|-1+\left(1-\dfrac{1}{1+\alpha h}\right)}{\dfrac{h}{1+\alpha h}}$$

$$=\mu(A)+\alpha$$

d) $\mu(A+B)=\lim\limits_{h\to 0+}\frac{\|I_n+h(A+B)\|-1}{h}$

$$=\lim_{h\to 0+}\frac{\|(I_n+2hA)+(I_n+2hB)\|-2}{2h}$$

$$\leqslant\lim_{h\to 0+}\left[\frac{\|I_n+2hA\|-1}{2h}+\frac{\|I_n+2hB\|-1}{2h}\right]$$

$$=\mu(A)+\mu(B)$$

e) 因为

$$\frac{\|I_n+hA\|-1}{h}+\frac{\|I_n-hA\|-1}{h}\geqslant\frac{2\|I_n\|-2}{h}=0$$

所以

$$-\|A\|=\frac{-h\|A\|+1-1}{h}\leqslant-\frac{\|I_n-hA\|-1}{h}$$

$$\leqslant\frac{\|I_n+hA\|-1}{h}\leqslant\frac{\|I_n\|+h\|A\|-1}{h}=\|A\|$$

因此(8)式成立；

f) 设 $e\in F^n$，$\|e\|=1$ 且 $Ae=\lambda e$，则

$$\frac{\|I_n+hA\|-1}{h}=\frac{\|I_n+hA\|\|e\|-1}{h}$$

$$\geqslant\frac{\|e+hAe\|-1}{h}=\frac{\|(1+h\lambda)e\|-1}{h}$$

$$=\frac{|1+h\lambda|-1}{h}=\frac{1+h\mathrm{Re}\lambda(A)+o(h)-1}{h}$$

$$=\mathrm{Re}\lambda(A)+\frac{o(h)}{h}$$

于是

$$\mu(A)\geqslant\mathrm{Re}\lambda(A)$$

并由此可知

$$-\mu(-A)\leqslant\mathrm{Re}\lambda(A)$$

g） 因为

$$\|Ax\| = \frac{\|(I_n - hA)x - x\|}{h}$$

$$\geqslant \frac{\|x\| - \|(I_n - hA)x\|}{h}$$

$$\geqslant \frac{\|x\| - \|I_n - hA\|\|x\|}{h}$$

$$= -\frac{(\|I_n - hA\| - 1)\|x\|}{h}$$

所以

$$\|Ax\| \geqslant -\mu(-A)\|x\|$$

又由此可得

$$\|Ax\| = \|-Ax\| \geqslant -\mu(A)\|x\|$$

综合以上结果知(9)式成立。证毕。

§4.5 矩阵特征值的估计

特征值的计算比较麻烦，但在某些科技问题中往往并不需知道特征值的准确值，而只需知道特征值分布在复平面中的位置即可。这样一来，特征值的估计就显得很必要了。

这里主要利用矩阵的范数及测度，给出特征值的一些估计式，最后介绍有名的圆盘定理。

定理1 若 $A \in F^{n\times n}$，$\|A\|$是算子范数，则

$a)\max\limits_i |\lambda_i(A)| \leqslant \|A\|, i \in \underline{n}$

$b) -\|A\| \leqslant -\mu(-A) \leqslant \text{Re}\lambda_i(A) \leqslant \mu(A) \leqslant \|A\|, i \in \underline{n}$

$c)\min\limits_i |\lambda_i(A)| \geqslant \max\{-\mu(-A), -\mu(A)\}, i \in \underline{n}$

证 a) 取 $\theta \neq x \in F^n$，且 $Ax = \lambda x$，则由

$$\|Ax\| \leqslant \|A\|\|x\|$$

有

$$\|\lambda\| \leqslant \|A\|$$

由于 λ 是 A 的任意特征值，于是

$$\max_i |\lambda_i(A)| \leqslant \|A\| \quad i \in \underline{n}$$

b) 由上节定理的 e)，f)即得；

c)由上节定理的 g)，取 $\theta \neq x_i \in F^n$，且

$$Ax_i = \lambda_i x_i \quad (i \in \underline{n}), \text{则}$$

$$|\lambda_i(A)| \|x_i\| \geqslant \max\{-\mu(-A), -\mu(A)\} \|x_i\|$$

于是

$$|\lambda_i(A)| \geqslant \max\{-\mu(-A), -\mu(A)\}$$

故

$$\min_i |\lambda_i(A)| \geqslant \max\{-\mu(-A), -\mu(A)\} \quad i \in \underline{n}$$

定理2 设 $A, B \in F^{n \times n}$，$\|A\|$、$\|B\|$为算子范数，则

a) $\max\limits_i \lambda_i(A+B) \leqslant \|A\| + \|B\|, i \in \underline{n}$

b) $-\mu(-A) - \mu(-B) \leqslant \mathrm{Re}\lambda_i(A+B) \leqslant \mu(A) + \mu(B)$,

$$i \in \underline{n}$$

证 a)由定理1知

$$\max_i |\lambda_i(A+B)| \leqslant \|A+B\| \leqslant \|A\| + \|B\|$$

b)由于

$$-\mu(-A) - \mu(-B) \leqslant -\mu(-A-B) \leqslant \mathrm{Re}\lambda_i(A+B) \leqslant \mu(A+B) \leqslant \mu(A) + \mu(B)$$

故

$$-\mu(-A) - \mu(-B) \leqslant \mathrm{Re}\lambda_i(A+B) \leqslant \mu(A) + \mu(B)$$

推论 a)当$\|A\| = \|A\|_2$、$\mu(A) = \mu_2(A)$时，有

$$\max_i |\lambda_i(A+B)| \leqslant \max_i \sqrt{\lambda_i(A^H A)}$$

$$+\max_i\sqrt{\lambda_i(B^HB)} \qquad i\in \underline{n}$$

$$\min_i\lambda_i\left(\frac{A^H+A}{2}\right)+\min_i\lambda_i\left(\frac{B^H+B}{2}\right)\leqslant \mathrm{Re}\lambda_i(A+B)$$

$$\leqslant \max_i\lambda_i\left(\frac{A^H+A}{2}\right)+\max_i\lambda_i\left(\frac{B^H+B}{2}\right) \qquad i\in \underline{n}$$

b)特别当 $A=A^H$、$B=B^H$ 时，有

$$\max_i|\lambda_i(A+B)|\leqslant\max_i|\lambda_i(A)|+\max_i|\lambda_i(B)| \quad i\in\underline{n}$$

$$\min_i\lambda_i(A)+\min_i\lambda_i(B)|\leqslant\mathrm{Re}\lambda_i(A+B)$$

$$\leqslant\max_i\lambda_i(A)+\max_i\lambda_i(B) \qquad i\in\underline{n}$$

由定理2，推论的结论是显然成立的。

定理3 设 $A,B\in F^{n\times n}$，$\|A\|$为算子范数，则

$$\max_i|\lambda_i(AB)|\leqslant\|AB\|(\leqslant\|A\|\|B\|)$$

推论 a)当$\|A\|=\|A\|_2$时，有

$$\max_i|\lambda_i(AB)|\leqslant\max_i\sqrt{\lambda_i(A^HA)}\max_i\sqrt{\lambda_i(B^HB)} \qquad i\in\underline{n}$$

b)当 $A=A^H$、$B=B^H$ 且$\|A\|=\|A\|_2$时，有

$$\max_i|\lambda_i(AB)|\leqslant\max_i|\lambda_i(A)|\max_i|\lambda_i(B)| \qquad i\in\underline{n}$$

定理4 对任意的埃尔米特矩阵 $A,B\in F^{n\times n}$，有

$$\lambda_{\min}(B)\lambda_i(A^2)\leqslant\lambda_i(ABA)\leqslant\lambda_{\max}(B)\lambda_i(A^2) \qquad i\in\underline{n}$$

其中 特征值按 $\lambda_{\max}=\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n=\lambda_{\min}$编号。

证 在等式

$$\lambda_{\max}(B)A^2=A(\lambda_{\max}(B)I-B)A+ABA$$

$$-\lambda_{\min}(B)A^2=A(B-\lambda_{\min}(B)I)A-ABA$$

中，$\lambda_{\max}(B)I-B$ 与 $B-\lambda_{\min}(B)I$ 都是半正定的，又由于 $A=A^H$、$B=B^H$，故

$$A(\lambda_{\max}(B)I - B)A、A(B - \lambda_{\min}(B)I)A$$

也是半正定的，可证当 $B \geqslant O$ 时

$$\lambda_i(A) \leqslant \lambda_i(A+B) \qquad i \in \underline{n}$$

（证略），故

$$\lambda_i(ABA) \leqslant \lambda_i(\lambda_{\max}(B)A^2) \qquad i \in \underline{n}$$

$$\lambda_i(-ABA) \leqslant \lambda_i(-\lambda_{\min}(B)A^2) \quad i \in \underline{n}$$

因此有

$$\lambda_{\min}(B)\lambda_i(A^2) \leqslant \lambda_i(ABA) \leqslant \lambda_{\max}(B)\lambda_i(A^2)$$

推论 对任意半正定矩阵 $A, B \in F^{n\times n}$，有

$$\lambda_{\min}(B)\lambda_{\max}(A) \leqslant \lambda_{\min}(B)\mathrm{tr}A \leqslant \mathrm{tr}AB \leqslant \lambda_{\max}(B)\mathrm{tr}A$$

证 由于 A、B 均半正定，故

$$\lambda_i(A) \geqslant 0 \qquad \lambda_i(B) \geqslant 0 \qquad i \in \underline{n}$$

且 $A^{\frac{1}{2}}$ 有意义，由定理4有

$$\lambda_{\min}(B)\lambda_i(A) \leqslant \lambda_i(A^{\frac{1}{2}}BA^{\frac{1}{2}}) \leqslant \lambda_{\max}(B)\lambda_i(A) \qquad i \in \underline{n}$$

以上 n 式相加，并利用 $\mathrm{tr}AB = \mathrm{tr}BA$ 得

$$\lambda_{\min}(B)\mathrm{tr}A \leqslant \mathrm{tr}AB \leqslant \lambda_{\max}(B)\mathrm{tr}A$$

而 $\lambda_{\min}(B)\lambda_{\max}(A) \leqslant \lambda_{\min}(B)\mathrm{tr}A$ 是显然的。证毕。

以上四个定理，用矩阵的范数和测度给出了特征值的估计式。下面在给出用矩阵本身的元素来确定它的特征值在复平面的分布区域。由于这些区域是由一些圆域组成的，故这个定理又称为**圆盘定理**。

定理5 盖尔斯高林(Gerschgorin)圆盘定理

设 $A = (a_{ij}) \in C_{n\times n}$，且

$$R_i = \left\{ z \,\middle|\, |z - a_{ii}| \leqslant \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \right\} \qquad i \in \underline{n}$$

$$S_j = \left\{ z \,\middle|\, |z - a_{jj}| \leqslant \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| \right\} \qquad j \in \underline{n}$$

则 A 的全部特征值都在 z 平面的区域

$$T = (\bigcup_{i=1}^{n} R_i) \cap (\bigcap_{j=1}^{n} S_j)$$

之内。

证　设 $\theta \neq x = (x_1, x_2, \cdots, x_n)^T \in C^n$，且 $Ax = \lambda x$，$\max_k |x_k| = |x_{i_0}| \neq 0 (i_0 \in \underline{n})$，由于

$$\sum_{j=1}^{n} a_{ij} x_j = \lambda x_i \qquad i \in \underline{n}$$

所以

$$|\lambda - a_{i_0 i_0}| \leqslant \sum_{\substack{j=1 \\ j \neq i_0}}^{n} |a_{i_0 j}| \frac{|x_j|}{|x_{i_0}|} \leqslant \sum_{\substack{j=1 \\ j \neq i_0}}^{n} |a_{i_0 j}|$$

即

$$\lambda \in R_{i_0} \qquad\qquad i_0 \in \underline{n}$$

因此 A 的特征值 λ 必位于区域 $\bigcup_{i=1}^{n} R_i$ 之内。

又由于 A 与 A^T 的特征值相同，故同理可知 A 的特征值也必然位于区域 $\bigcup_{j=1}^{n} S_j$ 之内。于是 A 的全部特征值必位于区域

$$T = (\bigcup_{i=1}^{n} R_i) \cap (\bigcup_{j=1}^{n} S_j)$$

之内。

例1　估计矩阵

$$A = \begin{bmatrix} 1 & 0.1 & 0.2 & 0.3 \\ 0.5 & 3 & 0.1 & 0.2 \\ 1 & 0.3 & -1 & 0.5 \\ 0.2 & -0.3 & -0.1 & -4 \end{bmatrix}$$

的特征值的范围。

解 由定理5有

$$R_1 = \{z \mid |z-1| \leqslant 0.6\}$$
$$R_2 = \{z \mid |z-3| \leqslant 0.8\}$$
$$R_3 = \{z \mid |z+1| \leqslant 1.8\}$$
$$R_4 = \{z \mid |z+4| \leqslant 0.6\}$$
$$S_1 = \{z \mid |z-1| \leqslant 1.7\}$$
$$S_2 = \{z \mid |z-3| \leqslant 0.7\}$$
$$S_3 = \{z \mid |z+1| \leqslant 0.4\}$$
$$S_4 = \{z \mid |z+4| \leqslant 1\}$$

在 z 平面上画出这八个圆域，然后再找出

$$T = (\bigcup_{i=1}^{4} R_i) \cap (\bigcup_{j=1}^{4} S_j)$$

即为所求矩阵 A 的特征值分布的区域。

注意：定理5表明矩阵 A 的特征值必位于 n 个圆域 R_1、R_2、…、R_n（或 S_1、S_2、…、S_n）之内，但这并不等于说正好一个圆域内有一个特征值，然而却可证"由 k 个圆组成的连通部分内，一定有 k 个特征值"。

例2 易知矩阵

$$\begin{pmatrix} 1 & -0.8 \\ 0.5 & 0 \end{pmatrix}$$

的特征值为

$$\lambda_{1,2} = \frac{1}{2} \pm \frac{\sqrt{0.6}}{2} i \qquad i = \sqrt{-1}$$

从而

$$|\lambda_1| = |\lambda_2| = \sqrt{0.4} \doteq 0.6$$

这里

$$R_1 = \{z \mid |z-1| \leqslant 0.8\}, R_2 = \{z \mid |z| \leqslant 0.5\}$$

$$S_1 = \{z \mid |z-1| \leqslant 0.5\}, S_2 = \{z \mid |z| \leqslant 0.8\}$$

均构成了连通域，但特征值并不分布在 $R_2(S_1)$ 之内。

定理5还可改进如定理6。

定理6 设 $A=(a_{ij})\in C^{n\times n}$，$\eta_i(i\in \underline{n})$是任一组正实数，令

$$\rho_1 = \max_i \frac{\sum_{j=1}^{n} |a_{ij}|\eta_j}{\eta_i}$$

$$\rho_2 = \max_j \eta_j \sum_{i=1}^{n} \frac{|a_{ij}|}{\eta_i}$$

则

$$\max_i |\lambda_i(A)| \leqslant \min(\rho_1, \rho_2)$$

证 令 $P=\mathrm{diag}(\eta_1,\eta_2,\cdots,\eta_n)$，则

$$P^{-1}AP = \begin{bmatrix} a_{11} & a_{12}\left(\frac{\eta_2}{\eta_1}\right) & \cdots & a_{1n}\left(\frac{\eta_n}{\eta_1}\right) \\ a_{21}\left(\frac{\eta_1}{\eta_2}\right) & a_{22} & \cdots & a_{2n}\left(\frac{\eta_n}{\eta_2}\right) \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}\left(\frac{\eta_1}{\eta_n}\right) & a_{n2}\left(\frac{\eta_2}{\eta_n}\right) & \cdots & a_{nn} \end{bmatrix}$$

与 A 具有相同的特征值。又

$$\|P^{-1}AP\|_\infty = \rho_1, \|P^{-1}AP\|_1 = \rho_2$$

由于

$$\max_i |\lambda_i(A)| \leqslant \|A\|$$

故

$$\max_i |\lambda_i(A)| \leqslant \min(\rho_1, \rho_2)$$

根据定理5及其注意事项，如果用定理5求出的某些圆域较大（表明精度差），则可适当选取正数组 $\eta_1,\eta_2,\cdots,\eta_n$，使得

$P^{-1}AP$ 中相应的圆域变小(此时可能使原在 A 中较小的圆域变大,这已无影响,因为已知该较小的圆域内有特征值。所说的"改进"即指此而言,对此,读者可在作习题中的第十题的过程中加深理解。

定理7 (Hermite 矩阵特征值的摄动定理)

设 $A=A^H, E=E^H\in C^{n\times n}, \tilde{A}=A+E, \alpha_i, \beta_i, \tilde{\alpha}_i, i\in \underline{n}$ 分别是 A、E、$\tilde{A}$ 的特征值,并按降序排列,则

$$\alpha_i+\beta_n\leqslant\tilde{\alpha}_i\leqslant\alpha_i+\beta_1$$

证略(详见《矩阵的扰动分析》,孙继广著,科学出版社)。

习题四

一、证明下述结论:

1. $\|x\|=[|2x_1-3x_2|^2+|x_2|^2]^{\frac{1}{2}}$ 是 C^2 上的向量范数,其中 $x=\begin{pmatrix}x_1\\x_2\end{pmatrix}\in C^2$

2. 设 $\|x\|$ 是 F^n 中的向量范数,$A\in F^{n\times n}$,则 $\|Ax\|$ 也是 F^n 中向量范数的充要条件为 A 是可逆矩阵。

二、证明:

1. $\|A\|_{m_2}=(\mathrm{tr}A^HA)^{\frac{1}{2}}$

2. $\|A\|_{m_2}$ 与 $\|x\|_2$ 是相容的;$\|A\|_{m_\infty}$ 与 $\|x\|_1$、$\|x\|_2$ 都是相容的。

3. $\|AB\|_{m_2}\leqslant\min(\|A\|_2\|B\|_{m_2}, \|A\|_{m_2}\|B\|_2)$

三、证明:

1. 若 $A\in F^{n\times r}$,且 $A^HA=I_r$,则

$$\|A\|_2=1, \qquad \|A\|_{m_2}=\sqrt{r}$$

2. 若$\|A\|$与$\|x\|$是相容的，则

$$\|I\| \geqslant 1$$

3. 若$\|\cdot\|$是算子范数，则

a) $\|I\|=1$

b) $\|A^{-1}\| \geqslant \|A\|^{-1}$

c) $\|A^{-1}\|^{-1}=\min\limits_{x\neq\theta}\dfrac{\|Ax\|}{\|x\|}$

四、设$\|A\|_v$，$\|A\|_\mu$是对应于两个向量范数$\|x\|_v$、$\|x\|_\mu=\|Bx\|_v$的算子范数，B可逆，则

$$\|A\|_\mu = \|BAB^{-1}\|_v$$

五、证明：

$$|\mu(A)-\mu(B)| \leqslant \max\{|\mu(A-B)|,|\mu(B-A)|\}$$

$$\forall\ A,B \in F^{n\times n}$$

六、设$A\in F_n^{n\times n}$，λ为A的特征值，则

$$\|A^{-1}\|_2^{-1} \leqslant |\lambda| \leqslant \|A\|_2$$

七、证明：

$$-\mu(jA) \leqslant Im\lambda_i(A) \leqslant \mu(-jA) \qquad j=\sqrt{-1}$$

八、若$\|A\|=\|A\|_2$，$\mu(A)=\mu_2(A)$，则

1. $\min\limits_i\lambda_i\left(\dfrac{A^H+A}{2}\right)\leqslant \mathrm{Re}\lambda_i(A)\leqslant\max\limits_i\lambda_i\left(\dfrac{A^H+A}{2}\right)$

2. $\min\limits_i\lambda_i\left(\dfrac{A-A^H}{2j}\right)\leqslant Im\lambda_i(A)\leqslant\max\limits_i\lambda_i\left(\dfrac{A-A^H}{2j}\right)$

$$j=\sqrt{-1}$$

3. 若$A=A^H$，则

$$-\max_i|\lambda_i(A)|\leqslant\min_i\lambda_i(A)\leqslant\lambda_i(A)\leqslant\max_i\lambda_i(A)$$

$$\leqslant\max_i|\lambda_i(A)| \qquad i\in\underline{n}$$

九、设$A\in F^{n\times n}$，若

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \qquad i \in \underline{n}$$

则

1. $\max_i \lambda_i(A) \leqslant \max_i 2|a_{ij}|$

2. $\lambda_i(A) \neq 0 \qquad i \in \underline{n}$

3. A 可逆。

十、利用§4.5定理5与定理6(取 $\eta_1=1$、$\eta_2=1$、$\eta_3=\frac{1}{10}$)分别确定矩阵

$$A = \begin{bmatrix} 0.9 & 0.01 & 0.12 \\ 0.01 & 0.8 & 0.13 \\ 0.01 & 0.02 & 0.4 \end{bmatrix}$$

的特征值所在的区域。

第五章　矩阵分析

在建立了范数概念之后，就可以讨论向量序列和矩阵序列的极限、矩阵级数、矩阵的微分等分析概念了，从而构成了矩阵分析的内容。

§5.1　向量序列和矩阵序列的极限

定义1　设 $x_n, x \in V(F, \|\ \|)$　$n=1,2,\cdots$，若

$$\|x_n - x\| \to 0 \qquad (n \to +\infty) \tag{1}$$

则称向量序列$\{x_n\}$收敛于向量 x，或说向量 x 是向量序列$\{x_n\}$当 $n \to +\infty$时的极限，可记为

$$\lim_{n \to +\infty} x_n = x \tag{2}$$

或

$$x_n \to x \qquad (n \to +\infty) \tag{3}$$

由向量范数之间的等价关系可知，在某一向量范数意义下收敛，在其它向量范数意义下也一定收敛。

例1　设 $x=(1,1,\cdots,1)^T \in R^n$，且

$$x_m = \left(1+\frac{1}{2^m}, 1+\frac{1}{3^m}, \cdots, 1+\frac{1}{(n+1)^m}\right)^T$$

则

$$\lim_{m \to +\infty} x_m = x$$

证　由于

$$\lim_{m\to+\infty}\|x_m - x\| = \lim_{m\to+\infty}\max_t \frac{1}{t^m}$$

$$= \lim_{m\to\infty}\frac{1}{2^m} = 0$$

所以

$$\lim_{m\to+\infty} x_m = x$$

例2 设

$$x_m = \sum_{k=0}^{m}\frac{t^k}{k!} \qquad t\in[0,1], m = 0,1,2,\cdots$$

是线性空间 $P[t]$的向量序列，又

$$\|x_m\|_\infty = \max_{t\in[0,1]}|x_m|$$

虽然 $n>m$ 时

$$\|x_m - x_n\|_\infty \leqslant \frac{1}{(m+1)!} + \frac{1}{(m+2)!} + \cdots + \frac{1}{n!}$$

只要 m、n 充分大，上面不等式右端可以任意地小，由微积分知

$$\lim_{m\to+\infty} x_m = e^t$$

虽然向量序列$\{x_m\}$的极限存在，但 e^t 并不是 $P[t]$中的向量，由此可见，在线性赋范空间中，柯西(Cauchy)收敛原理只是向量序列收敛的必要条件，即若$\{x_m\}$收敛，则对任给 $\varepsilon>0$，存在序号 N，使当 $m,n>N$ 时，恒有

$$\|x_m - x_n\| < \varepsilon$$

成立，但其逆不真。

若线性赋范空间中任一收敛向量序列的极限均属于此线性赋范空间，则称此线性赋范空间为**完备的线性赋范空间或巴拿赫**(Banach)**空间**，在巴拿赫空间中，柯西收敛原理成立。

如果线性赋范空间是 R^n 或 C^n，则它是完备的线性赋范

空间，柯西收敛原理对向量序列成立，且在向量范数意义下的收敛与各坐标序列的同时收敛等价。

定理1 $x_m=(x_1^{(m)},x_2^{(m)},\cdots,x_n^{(m)})^T$，$x=(x_1,x_2,\cdots,x_n)^T\in F^n$，则向量序列$\{x_m\}$收敛于$x$的充分条件为，坐标序列$\{x_i^{(m)}\}$收敛于$x_i$，$i\in \underline{n}$。

证 必要性：设$\|x\|=\|x\|_\infty=\max\limits_i|x_i|$，则

$$\|x_m-x\|_\infty\to 0 \qquad (m\to+\infty)$$

即由

$$\max_i|x_i^{(m)}-x_i|\geqslant|x_i^{(m)}-x_i|$$

知

$$\lim_{m\to+\infty}x_i^{(m)}=x_i \qquad i\in\underline{n}$$

充分性：若

$$\lim_{m\to+\infty}x_i^{(m)}=x_i \qquad i\in\underline{n}$$

则

$$\lim_{m\to+\infty}\max_i|x_i^{(m)}-x_i|=0$$

即

$$\|x_m-x\|\to 0 \qquad m\to+\infty$$

再由范数的等价性知，上述结论对F^n中任意向量范数均成立。

定理2 （柯西收敛原理）设$x_m\in F^n$，则向量序列$\{x_m\}$收敛的充要条件为：对任给$\varepsilon>0$，存在序号N，使当$m,p>N$时，恒有

$$\|x_m-x_p\|<\varepsilon$$

成立。

由定理1即知定理2的结论是显然成立的。

定义2 设$A_m=(a_{ij}^{(m)})$，$A=(a_{ij})\in C^{p\times n}$

若

$$\|A_m - A\| \to 0 \qquad (m \to +\infty) \tag{4}$$

则称矩阵序列$\{A_m\}$收敛于 A，或说 A 是矩阵序列$\{A_m\}$当 $m \to +\infty$时的极限。可记为

$$\lim_{m \to +\infty} A_m = A \tag{5}$$

或

$$A_m \to A \qquad (m \to +\infty) \tag{6}$$

由于矩阵序列亦可视为向量序列，因此，有与定理1、定理2相似的结论，这里不再重述。

关于矩阵序列的极限运算有下述定理：

定理3 设 $A_m, B_m, A, B \in F^{n \times n} (m=1,2,\cdots)$，且 $\lim_{m \to +\infty} A_m = A$，$\lim_{m \to +\infty} B_m = B$，则

a) $$\lim_{m \to +\infty} (A_m \pm B_m) = A \pm B \tag{7}$$

b) $$\lim_{m \to +\infty} A_m B_m = AB \tag{8}$$

c) $$\lim_{m \to +\infty} \lambda A_m = \lambda A \qquad \forall \lambda \in F \tag{9}$$

d) 若 A_m、A 均可逆，则

$$\lim_{m \to +\infty} A_m^{-1} = A^{-1} \tag{10}$$

证 a)只须注意

$$\|(A_m \pm B_m) - (A \pm B)\| \leqslant \|A_m - A\| + \|B_m - B\|$$

即知(7)式成立；

b) 由于

$$\|A_m B_m - AB\| = \|(A_m B_m - A_m B) + (A_m B - AB)\|$$
$$\leqslant \|A_m\| \|B_m - B\| + \|A_m - A\| \|B\|$$

当 m 充分大时，可使$\|A_m\| < \|A\| + 1$，故知当 $m \to +\infty$时(8)式成立；

c) 显然；

d) 记 adj$A$ 为 A 的伴随矩阵，则由

$$\det A_m \neq 0, \quad \det A \neq 0 \qquad m = 1,2,\cdots$$
$$\lim_{m\to+\infty} \det A_m = \det A$$
$$\lim_{m\to+\infty} \mathrm{adj} A_m = \mathrm{adj} A$$

知

$$\lim_{m\to+\infty} A_m^{-1} = \lim_{m\to+\infty} \frac{\mathrm{adj} A_m}{\det A_m} = \frac{\mathrm{adj} A}{\det A} = A^{-1}$$

证毕。

定理4 设 $A_m = A_m^T, P = P^T \in R^{n\times n}$，且

$$A_m \geqslant A_{m+1} \geqslant P \qquad m = 1,2,\cdots \tag{11}$$

则 $\lim\limits_{m\to+\infty} A_m$ 存在。

证 由假设知

$$x^T A_m x \geqslant x^T A_{m+1} x \geqslant x^T P x \qquad \forall\, x \in R^n \tag{12}$$

取

$$x = e_i = (0,\cdots,0,\underbrace{1}_{\text{第}i\text{个}},0,\cdots,0)^T \in R^n \quad i \in \underline{n}$$

代入(12)式得

$$a_{ii}^{(m)} \geqslant a_{ii}^{(m+1)} \geqslant p_{ii} \quad i \in \underline{n} \quad m = 1,2,\cdots \tag{13}$$

其中 $A_m = (a_{ij}^{(m)}), A_{m+1} = (a_{ij}^{(m+1)}), P = (p_{ij})$，故$\{a_{ii}^{(m)}\}$的极限存在($i \in \underline{n}$)，再取 $x = e_i + e_j$ 代入(12)式，得

$$a_{ij}^{(m)} + a_{ii}^{(m)} + a_{ji}^{(m)} + a_{jj}^{(m)} = a_{ii}^{(m)} + a_{jj}^{(m)} + 2a_{ij}^{(m)}$$
$$\geqslant a_{ii}^{(m+1)} + a_{jj}^{(m+1)} + 2a_{ij}^{(m+1)} \geqslant p_{ii} + p_{jj} + 2p_{ij}$$
$$i,j \in \underline{n} \quad m = 1,2,\cdots\cdots$$

由$\{a_{ii}^{(m)} + a_{jj}^{(m)} + 2a_{ij}^{(m)}\}$及$\{a_{ii}^{(m)}\}$收敛，知$\{a_{ij}^{(m)}\}$收敛，于是得知 $\lim\limits_{m\to+\infty} A_m$ 存在。

定理5 设 $A\in C^{n\times n}$，则 $\lim\limits_{m\to+\infty} A^m=O$ 的充分条件为

$$\max_i|\lambda_i(A)|<1$$

证 设 $T^{-1}AT=J$，其中 J 为 A 的约当标准形。

必要性：设 $\lim\limits_{m\to+\infty} A^m=O$，由于 $A^m=TJ^mT^{-1}$，故有 $\lim\limits_{m\to+\infty} J^m=O$。于是 $\lim\limits_{m\to+\infty} J_{ik}^m=O$，$i\in\underline{\sigma}$，$k\in\underline{\alpha_k}$。其中 J_{ik} 为 A 的约当块，σ 为 A 的相异特征值的个数，α_k 为特征值 λ_k 的几何重复度，故 $\lim\limits_{m\to+\infty} \lambda_i^m=0$，$i\in\underline{\sigma}$，所以 $\max\limits_i|\lambda_i(A)|<1$；

充分性：设 $\max\limits_i|\lambda_i(A)|<1$，则 $\lim\limits_{m\to+\infty} \lambda_i^m=0$，$i\in\underline{\sigma}$，由此可知，$\lim\limits_{m\to+\infty} J_{jk}^m=O$，从而有

$$\lim_{m\to+\infty} J^m=O$$

$$\lim_{m\to+\infty} A^m=O$$

证毕。

§5.2 矩阵级数

由于矩阵函数及系统微分方程的解都常常采用矩阵级数的形式表示，故有必要介绍矩阵级数的有关基本概念。

定义 设 $A_k\in F^{n\times n}(\|\cdot\|)$，$k=1,2,\cdots$，则称形式和

$$\sum_{k=1}^{+\infty}A_k=A_1+A_2+\cdots+A_k+\cdots \tag{1}$$

为矩阵级数，若 $A_k=(a_{ij}^{(k)})_{n\times n}$，且

$$\sum_{k=1}^{+\infty}a_{ij}^{(k)}(=a_{ij})\qquad i,j\in\underline{n} \tag{2}$$

收敛，则称**矩阵级数**(1)**收敛**，并且称 $A=(a_{ij})$ 为**矩阵级数**(1)**的和**，又若

$$\sum_{k=1}^{+\infty}\|A_k\|=\|A_1\|+\|A_2\|+\cdots+\|A_k\|+\cdots \tag{3}$$

收敛，则称**矩阵级数(1)是绝对收敛**的。

定理1 矩阵级数(1)收敛的充要条件为，由

$$S_k = A_1 + A_2 + \cdots + A_k$$

组成的矩阵序列$\{S_k\}$收敛。

定理1是显然成立的。

定理2 矩阵级数(1)绝对收敛的充要条件为

$$\sum_{k=1}^{+\infty} |a_{ij}^{(k)}| \qquad i,j \in \underline{n} \tag{4}$$

收敛。

证 必要性：设(3)收敛，则由

$$|a_{ij}^{(k)}| \leqslant \max_i \sum_{j=1}^{n} |a_{ij}^{(k)}| = \|A_k\|_\infty \qquad i,j \in \underline{n}$$

及正项级数的比较判别法、范数的等价性，知级数(4)收敛；

充分性：设级数(4)收敛，则由

$$b_{ij}^{(m)} = \sum_{k=1}^{m} |a_{ij}^{(k)}| \qquad i,j \in \underline{n}$$

组成的序列$\{b_{ij}^{(m)}\}$ $i,j\in \underline{n}$收敛。于是由柯西收敛原理知，对任给$\frac{\varepsilon}{n^2}>0$，存在序号$N$，使当$m>N$时，对一切自然数$p$，有

$$|b_{ij}^{(m+p)} - b_{ij}^{(m)}| < |a_{ij}^{(m+1)}| + \cdots + |a_{ij}^{(m+p)}| \leqslant \frac{\varepsilon}{n^2}$$

令 $u_m = \sum_{k=1}^{m} \|A_k\|_\infty$，则

$$\begin{aligned}|u_{m+p} - u_m| &= \|A_{m+1}\|_\infty + \cdots + \|A_{m+p}\|_\infty \\ &= \max_i \sum_{j=1}^{n} |a_{ij}^{(m+1)}| + \cdots + \max_i \sum_{i=1}^{n} |a_{ij}^{(m+p)}| \\ &\leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}[|a_{ij}^{(m+1)}| + \cdots + |a_{ij}^{(m+p)}|] < \varepsilon\end{aligned}$$

所以$\{u_m\}$收敛，即$\sum_{k=1}^{+\infty}\|A_k\|_\infty$收敛，于是(1)绝对收敛。

例1 设 $A\in F^{n\times n}$,则矩阵幂级数

$$I_n+A+\frac{A^2}{2!}+\cdots+\frac{A^k}{k!}+\cdots \tag{5}$$

绝对收敛。

证 因

$$\left\|\frac{A^k}{k!}\right\|\leqslant\frac{\|A\|^k}{k!}$$

且

$$\sum_{k=1}^{+\infty}\frac{\|A\|^k}{k!}+\|I_n\|=e^{\|A\|}-1+\|I_n\|$$

故级数(5)绝对收敛。

同理可知,矩阵幂级数

$$A-\frac{A^3}{3!}+\frac{A^5}{5!}-\cdots+(-1)^{k-1}\frac{A^{2k-1}}{(2k-1)!}+\cdots \tag{6}$$

$$I_n-\frac{A^2}{2!}+\frac{A^4}{4!}-\cdots+(-1)^k\frac{A^{2k}}{(2k)!}+\cdots \tag{7}$$

亦绝对收敛。仿微积分,分别定义矩阵幂级数(5)、(6)、(7)的和为矩阵函数 e^A、$\sin A$ 与 $\cos A$,即

$$e^A=I_n+A+\frac{A^2}{2!}+\cdots+\frac{A^k}{k!}+\cdots\cdots \tag{8}$$

$$\sin A=A-\frac{A^3}{3!}+\cdots+(-1)^{k-1}\frac{A^{2k-1}}{(2k-1)!}+\cdots\cdots \tag{9}$$

$$\cos A=I_n-\frac{A^2}{2!}+\cdots+(-1)^k\frac{A^{2k}}{(2k)!}+\cdots\cdots \tag{10}$$

例2 设 $A\in F^{n\times n}(\|\cdot\|)$,且 $\|A\|<1$,则矩阵幂级数

$$I_n+A+A^2+\cdots+A^k+\cdots \tag{11}$$

绝对收敛,且其和为矩阵 $(I_n-A)^{-1}$。

证 因 $\|A\|<1$,故知(11)绝对收敛,由于此时

$$\lim_{k\to+\infty} A^k = O \tag{12}$$

且由

$$\|(I_n - A)x\| \geqslant (1 - \|A\|)\|x\| > 0 \quad \forall\ \theta \neq x \in F^n$$

知

$$(I_n - A)x = \theta$$

只有零解，故 $\det(I_n - A) \neq 0$，即 $I_n - A$ 可逆。

若记

$$S_k = I_n + A + A^2 + \cdots + A^k$$

则

$$S_k(I_n - A) = I_n - A^{k+1}$$

由(12)式及$(I_n - A)^{-1}$存在知

$$\lim_{k\to+\infty} S^k = (I_n - A)^{-1}$$

关于矩阵级数的运算有下述性质。

定理3 设有两个矩阵级数

$$A_1 + A_2 + \cdots + A_k + \cdots \quad A_k \in F^{n\times n}(\|\cdot\|) \tag{13}$$

$$B_1 + B_2 + \cdots + B_k + \cdots \quad B_k \in F^{n\times n}(\|\cdot\|) \tag{14}$$

则

$a)$ 若级数(13)、(14)均收敛，且其和分别为 A 与 B，则级数

$$\sum_{k=1}^{+\infty}(A_k \pm B_k)、\quad \sum_{k=1}^{+\infty}\lambda A_k \qquad \lambda \in F$$

也收敛，且

$$\sum_{k=1}^{+\infty}(A_k \pm B_k) = A \pm B$$

$$\sum_{k=1}^{+\infty}\lambda A_k = \lambda A \qquad \forall\ \lambda \in F$$

$b)$ 若级数(13)、(14)均绝对收敛，且其和分别为 A 与

B，则其积

$$A_1B_1 + (A_1B_2 + A_2B_1) + \cdots + (A_1B_k + A_2B_{k-1} + \cdots + A_kB_1) + \cdots \tag{15}$$

也绝对收敛，且其和为 AB。

证 $a)$ 显然；

$b)$ 由于级数(13)、(14)均绝对收敛，故 $\|A_1\|\|B_1\| + \|A_1\|\|B_2\| + \|A_2\|\|B_1\| + \cdots + \|A_1\|\|B_k\| + \cdots + \|A_k\|\|B_1\| + \cdots$ 收敛，于是级数(15)绝对收敛。若以 S_{1k}、S_{2k}、S_{3k} 分别表示级数(13)、(14)、(15)的前 k 项和，$\tilde{S}_{1k}$、$\tilde{S}_{2k}$、$\tilde{S}_{3k}$ 分别表示级数(13)、(14)、(15)各项的范数所构成的正项级数的前 k 项和，于是由

$$\lim_{k\to+\infty} S_{1k}S_{2k} = AB$$

及

$$\|S_{1k}S_{2k} - S_{3k}\| \leqslant \tilde{S}_{1k}\tilde{S}_{2k} - \tilde{S}_{3k}$$

知

$$\lim_{k\to+\infty} S_{3k} = AB$$

例3 易证当 $A, B \in F^{n\times n}(\|\cdot\|)$，且 $AB = BA$ 时，

$$e^{A+B} = e^A e^B$$

此外

$$e^{jA} = \cos A + j\sin A \qquad j = \sqrt{-1}$$

$$\sin(A \pm B) = \sin A\cos B \pm \cos A\sin B$$

$$\cos(A \pm B) = \cos A\cos B \mp \sin A\sin B$$

§5.3 克罗内克(Kronecker)积

矩阵之间的克罗内克积是一种新的矩阵运算，用它研究矩阵方程，有时表示起来将十分简洁，研究矩阵微分运算时将要用到它，在此作简要介绍。

定义 设 $A=(a_{ij})\in F^{m\times n}$，$B\in F^{p\times q}$，则

$$A\otimes B=\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}_{mp\times nq}$$

称为矩阵 A 与矩阵 B 的**克罗内克积**。

例1 设

$$A=\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}、\quad B=\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & -1 \\ 2 & 0 & 0 & 1 \end{bmatrix}$$

则

$$A\otimes B=\begin{bmatrix} B & -B \\ 2B & B \end{bmatrix}=\begin{bmatrix} 1 & 0 & 2 & 0 & -1 & 0 & -2 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 2 & 0 & 0 & 1 & -2 & 0 & 0 & -1 \\ 2 & 0 & 4 & 0 & 1 & 0 & 2 & 0 \\ 0 & 2 & 0 & -2 & 0 & 1 & 0 & -1 \\ 4 & 0 & 0 & 2 & 2 & 0 & 0 & 1 \end{bmatrix}$$

矩阵的克罗内克积具有以下性质。

定理1 设 $A\in F^{m\times n}$、$B\in F^{p\times q}$、$C\in F^{r\times s}$，则

a) $\lambda(A\otimes B)=(\lambda A)\otimes B=A\otimes(\lambda B)\quad \forall\ \lambda\in F$

b) $A\otimes(B\otimes C)=(A\otimes B)\otimes C=A\otimes B\otimes C$

c) $(A\otimes B)^T=A^T\otimes B^T$，$(A\otimes B)^H=A^H\otimes B^H$

d) 当 $D\in F^{h\times k}$ 且 $nq=rk$ 时，

$$(A\otimes B)(C\otimes D)$$

有意义，且当 $n=r$，$q=k$ 时，有

$$(A\otimes B)(C\otimes D)=(AC)\otimes(BD)$$

e) 当 $p=m$、$q=n$ 时，有

$$(A+B)\otimes C = A\otimes C + B\otimes C$$
$$C\otimes (A+B) = C\otimes A + C\otimes B$$

f) 当 $p=m, q=n, G\in F^{r\times t}$时，有

$$(A+B)\otimes(C+G) = A\otimes C + A\otimes G + B\otimes C + B\otimes G$$

g) 当 $m=n, p=q$，且 A^{-1}、B^{-1}存在时，有

$$(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$$

h) 当 $m=n, p=q$ 时，有

$$\mathrm{tr}(A\otimes B) = \mathrm{tr}A\,\mathrm{tr}B$$

证 a)、b)、c)、e)、f)由克罗内克积的定义是容易证明的；

d) $(A\otimes B)(C\otimes D) = (a_{ij}B)(c_{ji}D)$

$$= \left(\sum_{j=1}^{n} a_{ij}c_{ji}BD\right) = ((AC)_{ij}BD)$$
$$= (AC)\otimes(BD);$$

g) 因由 d)有

$$(A\otimes B)(A^{-1}\otimes B^{-1}) = (AA^{-1})\otimes(BB^{-1})$$
$$I_n\otimes I_p = I_{np}$$

h) $\mathrm{tr}(A\otimes B) = \sum_{i=1}^{n} a_{ii}\sum_{j=1}^{n} b_{ij} = \mathrm{tr}A\,\mathrm{tr}B$

由定理1容易导出定理2。

定理2 a) 当 $A=A^T$、$B=B^T$ 时 $A\otimes B$ 也是对称矩阵；当 $A=A^H$、$B=B^H$ 时 $A\otimes B$，也是埃尔米特矩阵；

b) 当 $U\in F^{n\times n}$、$V\in F^{m\times m}$均为酉(正交)矩阵时，则 $U\otimes V$ 也是酉(正交)矩阵；

c) U、V 同 b)，$A\in F^{n\times n}$、$B\in F^{m\times m}$，且

$$U^HAU = R, V^HBV = R_1$$

若 R、R_1均为上三角阵，则

$$(U \otimes V)^H(A \otimes B)(U \otimes V) = R \otimes R_1$$

显然 $R \otimes R_1$ 也是上三角阵；

d） 若 $A \in F^{n\times n}$、$B \in F^{m\times m}$ 均为正规阵，则

$$A \otimes B \quad 、\quad B \otimes A$$

也均为正规阵，且若 $U_1 \in U^{n\times n}$、$U_2 \in U^{m\times m}$，使

$$U_1{}^H A U_1 = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \Lambda$$

$$U_2{}^H B U_2 = \mathrm{diag}(\mu_1, \mu_2, \cdots, \mu_n) = M$$

则

$$(U_1 \otimes U_2)^H(A \otimes B)(U_1 \otimes U_2) = \mathrm{diag}(\lambda_1 M, \lambda_2 M, \cdots, \lambda_n M)$$

$$(U_2 \otimes U_1)^H(B \otimes A)(U_2 \otimes U_1) = \mathrm{diag}(\mu_1 \Lambda, \mu_2 \Lambda, \cdots, \mu_n \Lambda)$$

e） 若 $A \in F^{m\times n}$、$B \in F^{p\times q}$ 的奇值分解分别为

$$A = U_1 D_1 V_1{}^H 、\quad B = U_2 D_2 V_2{}^H$$

A 与 B 的正奇值分别为 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 与 $\beta_1, \beta_2, \cdots, \beta_s$，则 $A \otimes B$ 的正奇值为 $\alpha_1\beta_1, \alpha_1\beta_2, \cdots, \alpha_1\beta_s, \alpha_2\beta_1, \cdots, \alpha_2\beta_s, \cdots, \alpha_r\beta_1, \cdots, \alpha_r\beta_s$，且

$$A \otimes B = (U_1 \otimes U_2)(D_1 \otimes D_2)(V_1 \otimes V_2)^H$$

是 $A \otimes B$ 的奇值分解；

f） 设

$$T^{-1}AT = \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

$$T_1{}^{-1}BT_1 = J = \mathrm{diag}(J_1, J_2, \cdots, J_n)$$

则

$$\begin{aligned}(T \otimes T_1)^{-1}(A \otimes B)(T \otimes T_1) &= \Lambda \otimes J \\ &= \mathrm{diag}(\lambda_1 J_1, \cdots, \lambda_1 J_\sigma, \lambda_2 J_1, \cdots, \lambda_2 J_\sigma, \cdots, \lambda_\sigma J_1, \\ &\quad \cdots, \lambda_\sigma J_\sigma)\end{aligned}$$

是 $A \otimes B$ 的约当标准形。

读者可选择其中一部分练习证明。

§5.4 矩阵的微分

矩阵微分的引进，为微分方程组的研究带来方便。因此，它是系统工程与控制论的基础，为了叙述简明起见，略去矩阵微分的范数形式的定义．而直接采用公式形式给出，并对不同情况的矩阵微分叙述其计算方法。

一、相对于数量变量的微分法

对于 n 维函数向量

$$\boldsymbol{a}(t) = (a_1(t), a_2(t), \cdots, a_n(t))^T$$

定义它对数量变量 t 的导数为

$$\frac{\mathrm{d}\boldsymbol{a}(t)}{\mathrm{d}t} \triangleq \left(\frac{\mathrm{d}a_1(t)}{\mathrm{d}t}, \frac{\mathrm{d}a_2(t)}{\mathrm{d}t}, \cdots, \frac{\mathrm{d}a_n(t)}{\mathrm{d}t}\right)^T$$

对于 $m\times n$ 阶函数矩阵

$$A(t) = (a_{ij}(t))_{m\times n}$$

定义它对数量变量 t 的导数为

$$\frac{\mathrm{d}A(t)}{\mathrm{d}t} \triangleq \left(\frac{\mathrm{d}a_{ij}(t)}{\mathrm{d}t}\right)_{m\times n}$$

显然，若将 n 维函数向量 $\boldsymbol{a}(t)$ 作为 $n\times 1$ 阶函数矩阵看待，它们对数量变量 t 的导数的定义是一致的。因此，下面仅就函数矩阵，讨论其对数量变量的导数的计算法则。

定理1 设 $A(t)$、$B(t)$、$A^{-1}(t)$ 及数量函数 $\lambda(t)$ 对数量变量 t 均可导，则

a) $\dfrac{\mathrm{d}(A \pm B)}{\mathrm{d}t} = \dfrac{\mathrm{d}A}{\mathrm{d}t} \pm \dfrac{\mathrm{d}B}{\mathrm{d}t} \qquad \forall A、B \in F^{m\times n}$

b) $\dfrac{\mathrm{d}\lambda A}{\mathrm{d}t} = \dfrac{\mathrm{d}\lambda}{\mathrm{d}t}A + \lambda\dfrac{\mathrm{d}A}{\mathrm{d}t} \qquad \forall A \in F^{m\times n}$

c) $\dfrac{\mathrm{d}AB}{\mathrm{d}t} = \dfrac{\mathrm{d}A}{\mathrm{d}t}B + A\dfrac{\mathrm{d}B}{\mathrm{d}t} \qquad \forall A \in F^{m\times n}, B \in F^{n\times p}$

d） $\dfrac{dA^{-1}}{dt} = -A^{-1}\dfrac{dA}{dt}A^{-1} \qquad \forall\, A \in F_n^{n\times n}$

证 a)、b)由导数定义即得；

c) 设 $\boldsymbol{a}(t)=(\alpha_1(t),\alpha_2(t),\cdots,\alpha_n(t))^T$，$\boldsymbol{b}(t)=(\beta_1(t),\beta_2(t),\cdots,\beta_n(t))^T$，则

$$\frac{d\boldsymbol{a}^T\boldsymbol{b}}{dt}=\sum_{i=1}^{n}\frac{d\alpha_i\beta_i}{dt}=\sum_{i=1}^{n}\left(\frac{d\alpha_i}{dt}\beta_i+\alpha_i\frac{d\beta_i}{dt}\right)$$

$$=\frac{d\boldsymbol{a}^T}{dt}\boldsymbol{b}+\boldsymbol{a}^T\frac{d\boldsymbol{b}}{dt}$$

设

$$A(t)=\begin{bmatrix}\boldsymbol{a}_1(t)^T\\ \boldsymbol{a}_2(t)^T\\ \vdots\\ \boldsymbol{a}_m(t)^T\end{bmatrix}$$

$$B(t)=(\boldsymbol{b}_1(t),\boldsymbol{b}_2(t),\cdots,\boldsymbol{b}_p(t))$$

其中 $\boldsymbol{a}_i(t)^T=(a_{i1}(t),a_{i2}(t),\cdots,a_{in}(t)) \quad i\in\underline{m}$

$\boldsymbol{b}_j(t)=(b_{1j}(t),b_{2j}(t),\cdots,b_{nj}(t))^T \quad j\in\underline{p}$

则由

$$AB=(\boldsymbol{a}_i{}^T\boldsymbol{b}_j)_{m\times p}$$

知

$$\frac{dAB}{dt}=\left(\frac{d\boldsymbol{a}_i{}^T\boldsymbol{b}_j}{dt}\right)_{m\times p}$$

$$=\left(\frac{d\boldsymbol{a}_i{}^T}{dt}\boldsymbol{b}_i+\boldsymbol{a}_i{}^T\frac{d\boldsymbol{b}_j}{dt}\right)_{m\times p}$$

$$=\frac{dA}{dt}B+A\frac{dB}{dt}$$

d) 由于 $AA^{-1}=I_n$，故

$$\frac{dAA^{-1}}{dt}=\frac{dA}{dt}A^{-1}+A\frac{dA^{-1}}{dt}=O$$

于是

$$\frac{\mathrm{d}A^{-1}}{\mathrm{d}t} = -A^{-1}\frac{\mathrm{d}A}{\mathrm{d}t}A^{-1}$$

例1 设 $A \in F^{n\times n}(\|\cdot\|)$是常数矩阵，则

a) $\dfrac{\mathrm{d}e^{At}}{\mathrm{d}t} = Ae^{At}$

b) $\dfrac{\mathrm{d}}{\mathrm{d}t}\sin At = A\cos At$

c) $\dfrac{\mathrm{d}}{\mathrm{d}t}\cos At = -A\sin At$

这只要注意到 e^{At}、$\sin At$、$\cos At$ 的级数表达式及函数矩阵的导数计算法则，即知上述结果是成立的。

例2 设 $A \in F^{n\times n}(\|\cdot\|)$是常数矩阵，$\boldsymbol{x}(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T$，则

a) e^{At}是微分方程

$$\dot{x} = Ax \qquad \left(\dot{x} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}\right)$$

的基本解矩阵，故 $\boldsymbol{x} = e^{At}$为此微分方程的解矩阵；

b) 若 A 可逆，则

$$\ddot{x} + A^2\boldsymbol{x} = \boldsymbol{\theta}$$

的通解为

$$\boldsymbol{x} = (\sin At)\boldsymbol{c}_1 + (\cos At)\boldsymbol{c}_2 \quad \forall\ \boldsymbol{c}_1, \boldsymbol{c}_2 \in F^n$$

证 a) 容易验证 $e^{At}\boldsymbol{c}$ 是 $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ 的解向量，若 $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ 即为初值条件；显然，$e^{A(t-t_0)}\boldsymbol{x}_0$为 $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ 满足 $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$的特解，故 e^{At}为 $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ 的基本解矩阵；

b) 容易验证$(\sin At)\boldsymbol{c}_1$、$(\cos At)\boldsymbol{c}_2$为 $\ddot{\boldsymbol{x}} + A^2\boldsymbol{x} = \boldsymbol{\theta}$ 的解向量，故 $\boldsymbol{x} = (\sin At)\boldsymbol{c}_1 + (\cos At)\boldsymbol{c}_2$为其通解；因为，只须取 $\boldsymbol{c}_1 = A^{-1}\boldsymbol{x}_1$、$\boldsymbol{c}_2 = \boldsymbol{x}_0$，即可得 $\ddot{\boldsymbol{x}} + A^2\boldsymbol{x} = \boldsymbol{\theta}$ 满足初值条件

$$\boldsymbol{x}(0) = \boldsymbol{x}_0, \quad \dot{\boldsymbol{x}}(0) = \boldsymbol{x}_1$$

的特解。

例3 求二次型 $x^T A x$ 对 t 的导数。

其中 $x=x(t)$ 是 n 维函数向量；

$A=A^T\in R^{n\times n}$ 是数字矩阵。

解

$$\frac{\mathrm{d}x^T A x}{\mathrm{d}t}=\frac{\mathrm{d}x^T}{\mathrm{d}t}Ax+x^T\frac{\mathrm{d}Ax}{\mathrm{d}t}$$

$$=\frac{\mathrm{d}x^T}{\mathrm{d}t}Ax+x^T\left(\frac{\mathrm{d}A}{\mathrm{d}t}x+A\frac{\mathrm{d}x}{\mathrm{d}t}\right)$$

$$=\frac{\mathrm{d}x^T}{\mathrm{d}t}Ax+x^T A\frac{\mathrm{d}x}{\mathrm{d}t}$$

$$=2\frac{\mathrm{d}x^T}{\mathrm{d}t}Ax\left(=2x^T A\frac{\mathrm{d}x}{\mathrm{d}t}\right)$$

特别，当 $A=I_n$ 时，

$$\frac{\mathrm{d}x^T x}{\mathrm{d}t}=2x^T\frac{\mathrm{d}x}{\mathrm{d}t}=2\frac{\mathrm{d}x^T}{\mathrm{d}t}x$$

二、相对于向量变量的微分法

1. 数量函数的导数

设

$$f(x)=g(x_1,x_2,\cdots,x_n)$$

是以向量 $x=(x_1,x_2,\cdots,x_n)^T$ 为自变量的数量函数，定义

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x}\triangleq\begin{bmatrix}\frac{\partial g}{\partial x_1}\\ \frac{\partial g}{\partial x_2}\\ \vdots\\ \frac{\partial g}{\partial x_n}\end{bmatrix}=\left(\frac{\partial}{\partial x_i}\right)_{n\times 1}\otimes g$$

显然，它是 R^3 中数量场 $u=u(x,y,z)$ 的梯度

$$\mathrm{grad}\,u=\nabla u=\left(\frac{\partial u}{\partial x},\frac{\partial u}{\partial y},\frac{\partial u}{\partial z}\right)^T$$

的概念的推广，故也可记为

$$\frac{\mathrm{d}f(x)}{\mathrm{d}\boldsymbol{x}} = \mathrm{grad}\ g(x_1, x_2, \cdots, x_n) = \nabla g(x_1, x_2, \cdots, x_n)$$

同理可定义

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x^T} \triangleq \left(\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}, \cdots, \frac{\partial g}{\partial x_n}\right)$$

显然，若还有向量 $\boldsymbol{x}$ 的数量函数

$$h(\boldsymbol{x}) = k(x_1, x_2, \cdots, x_n)$$

则下列导数法则成立：

$$\frac{\mathrm{d}[f(\boldsymbol{x}) \pm h(\boldsymbol{x})]}{\mathrm{d}\boldsymbol{x}} = \frac{\mathrm{d}f(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} \pm \frac{\mathrm{d}h(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}$$

$$\frac{\mathrm{d}f(\boldsymbol{x})h(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = \frac{\mathrm{d}f(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}h(\boldsymbol{x}) + f(\boldsymbol{x})\frac{\mathrm{d}h(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}$$

例4 求函数

$$f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x} = x_1^2 + x_2^2 + \cdots + x_n^2$$

对 $\boldsymbol{x}$ 的导数。

解 $\frac{\mathrm{d}f(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = (2x_1, 2x_2, \cdots, 2x_n)^T = 2\boldsymbol{x}$

2. 函数向量的导数

设 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$，且

$$a(x) = (a_1(x), a_2(x), \cdots, a_m(x))^T$$

是向量 $\boldsymbol{x}$ 的函数向量，定义

$$\frac{\mathrm{d}a(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}^T} \triangleq \begin{bmatrix} \frac{\partial a_1}{\partial x_1} & \frac{\partial a_1}{\partial x_2} & \cdots & \frac{\partial a_1}{\partial x_n} \\ \frac{\partial a_2}{\partial x_1} & \frac{\partial a_2}{\partial x_2} & \cdots & \frac{\partial a_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial a_m}{\partial x_1} & \frac{\partial a_m}{\partial x_2} & \cdots & \frac{\partial a_m}{\partial x_n} \end{bmatrix}_{m\times n}$$

$$\frac{\mathrm{d}a(x)^T}{\mathrm{d}x} \triangleq \left[\frac{\mathrm{d}a(x)}{\mathrm{d}x^T}\right]^T$$

定理2 设 $a(x), b(x) \in F^m, x \in F^n, \lambda(x) \in F$，则

a) $$\frac{\mathrm{d}(a \pm b)}{\mathrm{d}x^T} = \frac{\mathrm{d}a}{\mathrm{d}x^T} \pm \frac{\mathrm{d}b}{\mathrm{d}x^T}$$

b) $$\frac{\mathrm{d}(\lambda a)}{\mathrm{d}x^T} = \frac{\mathrm{d}\lambda}{\mathrm{d}x^T} a + \lambda \frac{\mathrm{d}a}{\mathrm{d}x^T}$$

c) $$\frac{\mathrm{d}(a^T b)}{\mathrm{d}x^T} = b^T \frac{\mathrm{d}a}{\mathrm{d}x^T} + a^T \frac{\mathrm{d}b}{\mathrm{d}x^T}$$

证 a)、b)由定义即知；

c) $$\frac{\mathrm{d}(a^T b)}{\mathrm{d}x^T} = \left[\frac{\partial(a^T b)}{\partial x_1}, \frac{\partial(a^T b)}{\partial x_2}, \cdots, \frac{\partial(a^T b)}{\partial x_n}\right]$$

$$= \left[\frac{\partial a^T}{\partial x_1} b + a^T \frac{\partial b}{\partial x_1}, \cdots, \frac{\partial a^T}{\partial x_n} b + a^T \frac{\partial b}{\partial x_n}\right]$$

$$= b^T \frac{\mathrm{d}a}{\mathrm{d}x^T} + a^T \frac{\mathrm{d}b}{\mathrm{d}x^T}$$

特别

$$\frac{\mathrm{d}x}{\mathrm{d}x^T} = \frac{\mathrm{d}x^T}{\mathrm{d}x} = I_m$$

将定理2中各公式转置，即可得另一组对向量 x 的导数公式。

例5 a) 求行向量 $x^T A$ 对 x 的导数；

b) 求列向量 Bx 对 x^T 的导数。

其中 $A \in F^{n \times m}$、$B \in F^{m \times n}$ 均为数字矩阵，$x \in F^n$。

解 a) 设 $A = (a_1, a_2, \cdots, a_m), a_i \in F^n, i \in \underline{m}$，

则

$$x^T A = (x^T a_1, x^T a_2, \cdots, x^T a_m)$$

于是

$$\frac{\mathrm{d}(x^T A)}{\mathrm{d}x} = \left[\frac{\mathrm{d}x^T a_1}{\mathrm{d}x}, \frac{\mathrm{d}x^T a_2}{\mathrm{d}x}, \cdots, \frac{\mathrm{d}x^T a_m}{\mathrm{d}x}\right]$$

$$=\left(\frac{\mathrm{d}\boldsymbol{x}^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{a}_1+\frac{\mathrm{d}\boldsymbol{a}_1{}^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{x},\cdots,\frac{\mathrm{d}\boldsymbol{x}^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{a}_m+\frac{\mathrm{d}\boldsymbol{a}_m{}^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{x}\right)$$

$$=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m)=A$$

b) 设

$$B=\begin{bmatrix}\boldsymbol{b}_1{}^T\\ \boldsymbol{b}_2{}^T\\ \vdots\\ \boldsymbol{b}_m{}^T\end{bmatrix}$$

则

$$B\boldsymbol{x}=\begin{bmatrix}\boldsymbol{b}_1{}^T\boldsymbol{x}\\ \boldsymbol{b}_2{}^T\boldsymbol{x}\\ \vdots\\ \boldsymbol{b}_m{}^T\boldsymbol{x}\end{bmatrix}$$

于是

$$\frac{\mathrm{d}(B\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}^T}=\begin{bmatrix}\dfrac{\mathrm{d}\boldsymbol{b}_1{}^T\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}^T}\\ \dfrac{\mathrm{d}\boldsymbol{b}_2{}^T\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}^T}\\ \vdots\\ \dfrac{\mathrm{d}\boldsymbol{b}_m{}^T\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}^T}\end{bmatrix}$$

$$=\begin{bmatrix}\boldsymbol{x}^T\dfrac{\mathrm{d}\boldsymbol{b}_1}{\mathrm{d}\boldsymbol{x}^T}+\boldsymbol{b}_1{}^T\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}^T}\\ \vdots\\ \vdots\\ \boldsymbol{x}^T\dfrac{\mathrm{d}\boldsymbol{b}_m}{\mathrm{d}\boldsymbol{x}^T}+\boldsymbol{b}_m{}^T\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}^T}\end{bmatrix}=\begin{bmatrix}\boldsymbol{b}_1{}^T\\ \vdots\\ \vdots\\ \vdots\\ \boldsymbol{b}_m{}^T\end{bmatrix}=B$$

例6 求二次型 $\boldsymbol{x}^TA\boldsymbol{x}$ 对 $\boldsymbol{x}$ 的导数。

解 $\frac{\mathrm{d}(\boldsymbol{x}^T A\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{x}^T}{\mathrm{d}\boldsymbol{x}}A\boldsymbol{x}+\frac{\mathrm{d}(A\boldsymbol{x})^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{x}$

$=A\boldsymbol{x}+\frac{\mathrm{d}\boldsymbol{x}^T A^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{x}=(A+A^T)\boldsymbol{x}$

若 $A=A^T$，则

$$\frac{\mathrm{d}\boldsymbol{x}^T A\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}} = 2A\boldsymbol{x}$$

例7 求数量函数 $\boldsymbol{p}^T A\boldsymbol{x}$ 对 $\boldsymbol{x}$ 的导数。

其中 $\boldsymbol{p}\in F^n, A\in F^{n\times n}$，它们的元均为常数。

解 $\frac{\mathrm{d}(\boldsymbol{p}^T A\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{p}^T}{\mathrm{d}\boldsymbol{x}}A\boldsymbol{x}+\frac{\mathrm{d}(A\boldsymbol{x})^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{p}=A^T\boldsymbol{p}$

3. 函数矩阵的导数

设矩阵 $A=(a_{ij})\in F^{m\times l}$的元 a_{ij}是向量 $\boldsymbol{x}\in F^n$ 的函数，即

$$A(\boldsymbol{x}) = (a_{ij}(\boldsymbol{x}))_{m\times l}$$

定义

$$\frac{\mathrm{d}(A\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} \triangleq \begin{bmatrix} \frac{\partial A}{\partial x_1} \\ \frac{\partial A}{\partial x_2} \\ \vdots \\ \frac{\partial A}{\partial x_n} \end{bmatrix}_{mn\times l}$$

$$\frac{\mathrm{d}(A\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}^T} \triangleq \left(\frac{\partial A}{\partial x_1},\frac{\partial A}{\partial x_2},\cdots,\frac{\partial A}{\partial x_n}\right)_{m\times ln}$$

函数矩阵对向量的微分法则如下。

定理3

$a)$ $\frac{\mathrm{d}[A(\boldsymbol{x})\pm B(\boldsymbol{x})]}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}A(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}\pm\frac{\mathrm{d}B(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}$

$b)$ $\frac{\mathrm{d}[\lambda(\boldsymbol{x})A(\boldsymbol{x})]}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}\lambda(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}\otimes A+\lambda(\boldsymbol{x})\frac{\mathrm{d}A(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}$

其中 $\lambda(\boldsymbol{x})$是向量 $\boldsymbol{x}$ 的数量函数。

$c)\quad \dfrac{\mathrm{d}[A(\boldsymbol{x})B(\boldsymbol{x})]}{\mathrm{d}\boldsymbol{x}}=\dfrac{\mathrm{d}A(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}B(\boldsymbol{x})+[I_n\otimes A]\dfrac{\mathrm{d}B(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}$

证 $a)$ 由定义即知；

$$b)\quad \frac{\mathrm{d}[\lambda(\boldsymbol{x})A(\boldsymbol{x})]}{\mathrm{d}\boldsymbol{x}}=\begin{bmatrix}\dfrac{\partial\lambda A}{\partial x_1}\\ \dfrac{\partial\lambda A}{\partial x_2}\\ \vdots\\ \dfrac{\partial\lambda A}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\dfrac{\partial\lambda}{\partial x_1}A+\lambda\dfrac{\partial A}{\partial x_1}\\ \dfrac{\partial\lambda}{\partial x_2}A+\lambda\dfrac{\partial A}{\partial x_2}\\ \vdots\\ \dfrac{\partial\lambda}{\partial x_n}A+\lambda\dfrac{\partial A}{\partial x_n}\end{bmatrix}$$

$$=\begin{bmatrix}\dfrac{\partial\lambda}{\partial x_1}A\\ \dfrac{\partial\lambda}{\partial x_2}A\\ \vdots\\ \dfrac{\partial\lambda}{\partial x_n}A\end{bmatrix}+\begin{bmatrix}\lambda\dfrac{\partial A}{\partial x_1}\\ \lambda\dfrac{\partial A}{\partial x_2}\\ \vdots\\ \lambda\dfrac{\partial A}{\partial x_n}\end{bmatrix}=\frac{\mathrm{d}\lambda(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}\otimes A+\lambda(\boldsymbol{x})\frac{\mathrm{d}A(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}$$

$$c)\quad \frac{\mathrm{d}[A(\boldsymbol{x})B(\boldsymbol{x})]}{\mathrm{d}\boldsymbol{x}}=\begin{bmatrix}\dfrac{\partial AB}{\partial x_1}\\ \vdots\\ \dfrac{\partial AB}{\partial x_n}\end{bmatrix}$$

$$=\begin{bmatrix}\dfrac{\partial A}{\partial x_1}B+A\dfrac{\partial B}{\partial x_1}\\ \vdots\\ \dfrac{\partial A}{\partial x_n}B+A\dfrac{\partial B}{\partial x_n}\end{bmatrix}$$

$$=\frac{\mathrm{d}A}{\mathrm{d}\boldsymbol{x}}B+[I_n\otimes A]\frac{\mathrm{d}B}{\mathrm{d}\boldsymbol{x}}$$

例8 求 $\boldsymbol{x}^TA$ 对向量 $\boldsymbol{x}$ 的导数。

其中　A 为常数矩阵。

解　$\frac{\mathrm{d}\boldsymbol{x}^T A}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{x}^T}{\mathrm{d}\boldsymbol{x}}A+(I_n\otimes\boldsymbol{x}^T)\frac{\mathrm{d}A}{\mathrm{d}\boldsymbol{x}}=A$

例9　求 $\boldsymbol{p}^T A\boldsymbol{q}$ 对向量 $\boldsymbol{x}$ 的导数。

其中　$\boldsymbol{p}$、$\boldsymbol{q}$ 均为常向量；

A 为向量 $\boldsymbol{x}$ 的函数矩阵。

解　
$$\frac{\mathrm{d}[\boldsymbol{p}^T A\boldsymbol{q}]}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{p}^T}{\mathrm{d}\boldsymbol{x}}A\boldsymbol{q}+\frac{\mathrm{d}[A\boldsymbol{q}]^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{p}$$
$$=\frac{\mathrm{d}\boldsymbol{q}^T}{\mathrm{d}\boldsymbol{x}}A^T\boldsymbol{p}+[I_n\otimes\boldsymbol{q}^T]\frac{\mathrm{d}A^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{p}$$
$$=[I_n\otimes\boldsymbol{q}^T]\frac{\mathrm{d}A^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{p}$$

三、相对于矩阵的微分法

1. 数量矩阵函数的导数

设 $A\in F^{m\times l}$，$f(A)$ 为矩阵 A 的数量函数，定义

$$\frac{\mathrm{d}f(A)}{\mathrm{d}A}\triangleq\left[\frac{\partial f}{\partial a_{ij}}\right]_{m\times l}\qquad A=(a_{ij})_{m\times l}$$

显然，由此有

$$\frac{\mathrm{d}[f(A)\pm g(A)]}{\mathrm{d}A}=\frac{\mathrm{d}f(A)}{\mathrm{d}A}\pm\frac{\mathrm{d}g(A)}{\mathrm{d}A}$$
$$\frac{\mathrm{d}f(A)g(A)}{\mathrm{d}A}=\frac{\mathrm{d}f(A)}{\mathrm{d}A}g(A)+f(A)\frac{\mathrm{d}g(A)}{\mathrm{d}A}$$

例10　求二次型 $\boldsymbol{x}^T A\boldsymbol{x}$ 对 A 的导数。

其中　$A=(a_{ij})\in F^{n\times n}$，$\boldsymbol{x}\in F^n$ 与 A 无关。

解　$\frac{\mathrm{d}\boldsymbol{x}^T A\boldsymbol{x}}{\mathrm{d}A}=\left[\frac{\partial\boldsymbol{x}^T A\boldsymbol{x}}{\partial a_{ij}}\right]_{n\times n}=(x_i x_j)_{n\times n}=\boldsymbol{x}\boldsymbol{x}^T$

这里用到 $\boldsymbol{x}^T A\boldsymbol{x}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_i x_j$。

2. 向量矩阵函数的导数

设 $\boldsymbol{z}(A)=(z_1(A),z_2(A),\cdots,z_n(A))^T$，$A=(a_{ij})_{m\times l}$，定义

$$\frac{\mathrm{d}z(A)}{\mathrm{d}A} \triangleq \left(\frac{\partial z(A)}{\partial a_{ij}}\right)_{m\times l}$$

其中 $$\frac{\partial z(A)}{\partial a_{ij}} = \begin{bmatrix} \frac{\partial z_1(A)}{\partial a_{ij}} \\ \vdots \\ \frac{\partial z_n(A)}{\partial a_{ij}} \end{bmatrix}$$

仿此，可定义 $z(A)^T=(z_1(A),z_2(A),\cdots,z_n(A))$ 对矩阵 A 的导数

$$\frac{\mathrm{d}z(A)^T}{\mathrm{d}A} \triangleq \left(\frac{\partial z(A)^T}{\partial a_{ij}}\right)_{m\times nl}$$

其中 $$\frac{\partial z(A)^T}{\partial a_{ij}} = \left(\frac{\partial z_1(A)}{\partial a_{ij}},\cdots,\frac{\partial z_n(A)}{\partial a_{ij}}\right)$$

3. 矩阵函数的导数

设矩阵 G 是矩阵 A 的函数，即

$$G(A) = \begin{bmatrix} g_{11}(A) & g_{12}(A)\cdots g_{1q}(A) \\ g_{21}(A) & g_{22}(A)\cdots g_{2q}(A) \\ \cdots\ \cdots & \cdots\ \cdots\ \cdots \\ g_{p1}(A) & g_{p2}(A)\cdots g_{pq}(A) \end{bmatrix}$$

是以 $A\in F^{m\times n}$ 为自变量的 $p\times q$ 阶矩阵，定义

$$\frac{\mathrm{d}G(A)}{\mathrm{d}A} \triangleq \left(\frac{\partial G}{\partial a_{ij}}\right)_{pm\times qn} \doteq \nabla \otimes G$$

其中 $$\frac{\partial G}{\partial a_{ij}} = \begin{bmatrix} \frac{\partial g_{11}}{\partial a_{ij}}\cdots\frac{\partial g_{1q}}{\partial a_{ij}} \\ \cdots\ \ \cdots\ \ \cdots \\ \frac{\partial g_{p1}}{\partial a_{ij}}\cdots\frac{\partial g_{pq}}{\partial a_{ij}} \end{bmatrix}$$

与
$$\dot{\nabla}=\begin{bmatrix}\frac{\partial}{\partial a_{11}}\cdots\frac{\partial}{\partial a_{1n}}\\ \cdots\quad\cdots\quad\cdots\\ \frac{\partial}{\partial a_{m1}}\cdots\frac{\partial}{\partial a_{mn}}\end{bmatrix}$$

为哈米尔顿(Hamilton)算子矩阵。

定理4 *a*) $\frac{\mathrm{d}[G(A)\pm D(A)]}{\mathrm{d}A}=\frac{\mathrm{d}G(A)}{\mathrm{d}A}\pm\frac{\mathrm{d}D(A)}{\mathrm{d}A}$

b) 设 $\lambda(A)\in F$，则

$$\frac{\mathrm{d}[\lambda(A)G(A)]}{\mathrm{d}A}=\frac{\mathrm{d}\lambda(A)}{\mathrm{d}A}\otimes G(A)+\lambda(A)\frac{\mathrm{d}G(A)}{\mathrm{d}A}$$

c) $$\frac{\mathrm{d}[G(A)H(A)]}{\mathrm{d}A}=\frac{\mathrm{d}G(A)}{\mathrm{d}A}[I_n\otimes H(A)]+[I_m\otimes G(A)]\frac{\mathrm{d}H(A)}{\mathrm{d}A}$$

其中 $A\in F^{m\times n}$。

证明留给读者练习。

例11 求 y^TA^TAy 对 A 的导数。

其中 $A\in F^{m\times n}$、$y\in F^n$ 是常向量。

解
$$\begin{aligned}\frac{\mathrm{d}[y^TA^TAy]}{\mathrm{d}A}&=\left[\frac{\partial[y^TA^TAy]}{\partial a_{ij}}\right]\\&=\left[\frac{\partial[y^TA^T]}{\partial a_{ij}}Ay+y^TA^T\frac{\partial[Ay]}{\partial a_{ij}}\right]\\&=\left[y^T\frac{\partial A^T}{\partial a_{ij}}Ay+y^TA^T\frac{\partial A}{\partial a_{ij}}y\right]\\&=\left[\sum_{k=1}^{n}a_{ik}y_ky_j\right]+\left[\sum_{k=1}^{n}a_{ki}y_ky_j\right]\\&=Ayy^T+Ayy^T=2Ayy^T\end{aligned}$$

由此可得

$$\frac{\mathrm{d}(Ay-x)^T(Ay-x)}{\mathrm{d}A}=2(Ay-x)y^T$$

事实上

$$(Ay-x)^T(Ay-x)=y^TA^TAy-x^TAy-y^TA^Tx+x^Tx$$

由例10可知

$$\frac{\mathrm{d}x^TAy}{\mathrm{d}A}=xy^T$$

又因 $y^TA^Tx=(x^TAy)^T$，故$\dfrac{\mathrm{d}x^TAy}{\mathrm{d}A}=xy^T$，因此有上述结果。

例12 求 $\mathrm{tr}(Ay-x)(Ay-x)^T$、$\mathrm{tr}AB$、$\mathrm{tr}ACA^T$ 对 A 的导数。

其中 B、C 是常矩阵；

x、y 是常向量。

解 由于 $\mathrm{tr}(Ay-x)(Ay-x)^T=\mathrm{tr}(Ay-x)^T(Ay-x)$，故由例11知

$$\frac{\mathrm{d}\,\mathrm{tr}(Ay-x)(Ay-x)^T}{\mathrm{d}A}=2(Ay-x)y^T$$

又

$$\mathrm{tr}AB=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}b_{ji}$$

故

$$\frac{\mathrm{d}\,\mathrm{tr}AB}{\mathrm{d}A}=\left(\frac{\partial\,\mathrm{tr}AB}{\partial a_{kl}}\right)=(b_{lk})=B^T$$

其中 $A\in F^{n\times m}$、$B\in F^{m\times n}$。

由矩阵迹的性质有

$$\mathrm{tr}AB=\mathrm{tr}BA=\mathrm{tr}A^TB^T=\mathrm{tr}B^TA^T$$

故

$$\frac{\mathrm{d}\,\mathrm{tr}AB}{\mathrm{d}A}=\frac{\mathrm{d}\,\mathrm{tr}BA}{\mathrm{d}A}=\frac{\mathrm{d}\,\mathrm{tr}A^TB^T}{\mathrm{d}A}=\frac{\mathrm{d}\,\mathrm{tr}B^TA^T}{\mathrm{d}A}=B^T$$

又

$$\frac{\mathrm{dtr}ACA^T}{\mathrm{d}A}=\frac{\mathrm{dtr}AB_1}{\mathrm{d}A}+\frac{\mathrm{dtr}B_2A^T}{\mathrm{d}A}=B_1{}^T+B_2=A(C+C^T)$$

其中 $B_1=CA^T$、$B_2=AC$ 在求导时均看作常数矩阵。

例13 求 $f=(x-a-Bz)^T(x-a-Bz)$ 对矩阵 B 的导数。其中 $\boldsymbol{x}$、$\boldsymbol{a}$、$\boldsymbol{z}$ 均为常向量。

解 由例11知

$$\frac{\mathrm{d}(x-a-Bz)^T(x-a-Bz)}{\mathrm{d}B}$$

$$=\frac{\mathrm{d}(Bz-x+a)^T(Bz-x+a)}{\mathrm{d}B}=2(\boldsymbol{Bz}-\boldsymbol{x}+\boldsymbol{a})\boldsymbol{z}^T$$

四、复合函数微分法

1. 数量函数的求导公式

设 $f=f(\boldsymbol{y})$，$\boldsymbol{y}=\boldsymbol{y}(t)$，若 f、t 为数量变量，$\boldsymbol{y}$ 为向量变量，则

$$\frac{\mathrm{d}f}{\mathrm{d}t}=\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{y}^T}\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t}=\frac{\mathrm{d}\boldsymbol{y}^T}{\mathrm{d}t}\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{y}} \tag{1}$$

设 $f=f(\boldsymbol{y})$，$\boldsymbol{y}=\boldsymbol{y}(\boldsymbol{x})$，若 f 为数量变量，$\boldsymbol{x}$、$\boldsymbol{y}$ 为向量变量，则

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{y}^T}{\mathrm{d}\boldsymbol{x}}\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{y}} \tag{2}$$

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}^T}=\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{y}^T}\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}^T} \tag{3}$$

这只要注意到

$$\mathrm{d}f=\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{y}^T}\mathrm{d}\boldsymbol{y},\ \mathrm{d}\boldsymbol{y}=\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}^T}\mathrm{d}\boldsymbol{x}$$

即可。

又设 $f=f(\boldsymbol{x},\boldsymbol{y})$，$\boldsymbol{y}=\boldsymbol{y}(\boldsymbol{x})$，若 f 是常量变量，$\boldsymbol{x}$、$\boldsymbol{y}$ 是向量

变量，则

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\frac{\partial f}{\partial \boldsymbol{x}}+\frac{\mathrm{d}\boldsymbol{y}^T}{\mathrm{d}\boldsymbol{x}}\frac{\partial f}{\partial \boldsymbol{y}} \tag{4}$$

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}^T}=\frac{\partial f}{\partial \boldsymbol{x}^T}+\frac{\partial f}{\partial \boldsymbol{y}^T}\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}^T} \tag{5}$$

以上公式的证明略去。下面举例说明其应用。

例14 求 $f=\boldsymbol{x}^T A\boldsymbol{x}$ 对 $\boldsymbol{x}$ 的导数。

解 设 $\boldsymbol{y}=A\boldsymbol{x}$，于是由公式(4)

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\boldsymbol{y}+A^T\boldsymbol{x}=(A+A^T)\boldsymbol{x}$$

例15 求 $f=(\boldsymbol{x}-\boldsymbol{a}-B\boldsymbol{z})^T(\boldsymbol{x}-\boldsymbol{a}-B\boldsymbol{z})$对 $\boldsymbol{a}$ 的导数。

解 令 $\boldsymbol{y}=\boldsymbol{x}-\boldsymbol{a}-B\boldsymbol{z}$，则 $f=\boldsymbol{y}^T\boldsymbol{y}$，于是由公式(2)有

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{a}}=\frac{\mathrm{d}\boldsymbol{y}^T}{\mathrm{d}\boldsymbol{a}}\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{y}}=-I(2\boldsymbol{y})=-2(\boldsymbol{x}-\boldsymbol{a}-B\boldsymbol{z})$$

2. 向量函数的求导公式

设 $\boldsymbol{z}=\boldsymbol{z}(\boldsymbol{y})$，$\boldsymbol{y}=\boldsymbol{y}(t)$，若 t 为数量变量，$\boldsymbol{y}$、$\boldsymbol{z}$ 为向量变量，则

$$\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}t}=\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}\boldsymbol{y}^T}\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t} \tag{6}$$

设 $\boldsymbol{z}=(\boldsymbol{z}\boldsymbol{y})$，$\boldsymbol{y}=\boldsymbol{y}(\boldsymbol{x})$，若 $\boldsymbol{x}$、$\boldsymbol{y}$、$\boldsymbol{z}$ 均为向量变量，则

$$\frac{\mathrm{d}\boldsymbol{z}^T}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{y}^T}{\mathrm{d}\boldsymbol{x}}\frac{\mathrm{d}\boldsymbol{z}^T}{\mathrm{d}\boldsymbol{y}} \tag{7}$$

$$\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}\boldsymbol{x}^T}=\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}\boldsymbol{y}^T}\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}^T} \tag{8}$$

设 $\boldsymbol{z}=f(\boldsymbol{x},\boldsymbol{y})$，$\boldsymbol{y}=\boldsymbol{y}(\boldsymbol{x})$，若 $\boldsymbol{x}$、$\boldsymbol{y}$、$\boldsymbol{z}$ 均为向量变量，则

$$\frac{\mathrm{d}\boldsymbol{z}^T}{\mathrm{d}\boldsymbol{x}}=\frac{\partial f}{\partial \boldsymbol{x}}+\frac{\mathrm{d}\boldsymbol{y}^T}{\mathrm{d}\boldsymbol{x}}\frac{\partial f}{\partial \boldsymbol{y}} \tag{9}$$

$$\frac{\mathrm{d}z}{\mathrm{d}\boldsymbol{x}^T} = \frac{\partial f}{\partial \boldsymbol{x}^T} + \frac{\partial f}{\partial \boldsymbol{y}^T}\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}^T} \tag{10}$$

例16 求向量函数$[\sin(c^T\boldsymbol{x})]\boldsymbol{x}^T$对$\boldsymbol{x}$的导数。

其中 $\boldsymbol{c}$为常向量。

解 令$\boldsymbol{c}^T\boldsymbol{x}=y$，则$\boldsymbol{z}^T=(\sin y)\boldsymbol{x}^T$，由公式(9)有

$$\frac{\mathrm{d}\boldsymbol{z}^T}{\mathrm{d}\boldsymbol{x}} = \frac{\partial f}{\partial \boldsymbol{x}} + \frac{\mathrm{d}y^T}{\mathrm{d}\boldsymbol{x}}\frac{\partial f}{\partial y} = \sin y\frac{\mathrm{d}\boldsymbol{x}^T}{\mathrm{d}\boldsymbol{x}} + \frac{\mathrm{d}\boldsymbol{x}^T}{\mathrm{d}\boldsymbol{x}}\boldsymbol{c}(\cos y)\boldsymbol{x}^T$$

$$=(\sin y)I+\boldsymbol{c}\cdot(\cos y)\boldsymbol{x}^T$$

$$=[\sin(\boldsymbol{c}^T\boldsymbol{x})]I+[\cos(\boldsymbol{c}^T\boldsymbol{x})]\boldsymbol{c}\boldsymbol{x}^T$$

例17 求$f=(A\boldsymbol{x}-\boldsymbol{b})^TR(A\boldsymbol{x}-\boldsymbol{b})$对$\boldsymbol{x}$的导数。

其中 A、R为常数矩阵，$\boldsymbol{b}$是常数向量。

解 令$\boldsymbol{y}=A\boldsymbol{x}-\boldsymbol{b}$，则$f=\boldsymbol{y}^TR\boldsymbol{y}$，由公式(4)有

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{y}^T}{\mathrm{d}\boldsymbol{x}}\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{y}}=A^T(R\boldsymbol{y}+R^T\boldsymbol{y})$$

$$=A^T(R+R^T)(A\boldsymbol{x}-\boldsymbol{b})$$

§5.5 矩阵的积分

设$A(t)=(a_{ij}(t))_{m\times n}$，若$a_{ij}(t)$在$[t_1,t_2]$上可积，定义

$$\int_{t_1}^{t_2}A(t)\mathrm{d}t \triangleq \left(\int_{t_1}^{t_2}a_{ij}(t)\mathrm{d}t\right)_{m\times n}$$

$$\mathrm{d}A(t) \triangleq (\mathrm{d}a_{ij}(t))_{m\times n}$$

于是

$$\int_{t_1}^{t_2}[A(t)\pm B(t)]\mathrm{d}t = \int_{t_1}^{t_2}A(t)\mathrm{d}t \pm \int_{t_1}^{t_2}B(t)\mathrm{d}t$$

$$\int_{t_1}^{t_2}kA(t)\mathrm{d}t = k\int_{t_1}^{t_2}A(t)\mathrm{d}t \qquad k=\mathrm{const}$$

$$\mathrm{d}[A(t) \pm B(t)] = \mathrm{d}A(t) \pm \mathrm{d}B(t)$$

$\mathrm{d}[\lambda(t)A(t)] = (\mathrm{d}\lambda(t))A(t) + \lambda(t)\mathrm{d}A(t)$（其中 $\lambda(t)$ 为数量函数）

$$\mathrm{d}[A(t)B(t)] = (\mathrm{d}A(t))B(t) + A(t)\mathrm{d}B(t)$$

$$\mathrm{d}[A(t) \otimes B(t)] = (\mathrm{d}A(t)) \otimes B(t) + A(t) \otimes (\mathrm{d}B(t))$$

$$\mathrm{d}A^{-1}(t) = -A^{-1}(t)(\mathrm{d}A(t))A^{-1}(t)$$

以上性质，由函数矩阵的积分和微分的定义知是显然成立的。

下面讨论函数向量组的线性相关性问题。

定义1 设 $f_i(t) = (f_{i1}(t), f_{i2}(t), \cdots, f_{im}(t))^T$ $(i \in \underline{n})$ 是区间 $[t_1, t_2]$ 上的 n 个 m 维函数向量组成的函数向量组，若存在 n 个不全为零的常数 $a_i (i \in \underline{n})$，使得

$$\sum_{i=1}^{m} a_i f_i(t) \equiv \theta \qquad t \in [t_1, t_2]$$

则称这个函数向量组是**线性相关的**，否则称为是**线性无关的**。

注意：这里向量的分量是函数，因此，它与数 a_i 不在同一域内，故由向量组成的矩阵的秩来判断该向量组是否是线性相关的方法，在此已不再适用，为此引进**克兰姆（Gram）行列式**的概念。

定义2 设 $f_i(t) = (f_{i1}(t), f_{i2}(t), \cdots, f_{im}(t))^T$ $(i \in \underline{n})$ 是区间 $[t_1, t_2]$ 上的 n 个 m 维函数向量组成的函数向量组，记

$$g_{ij} = \int_{t_1}^{t_2} f_i^T(t) f_j(t) \mathrm{d}t \qquad i, j \in \underline{n}$$

则称

$$G = (g_{ij})_{n \times n} \in R^{n \times n}$$

为该函数向量组的克兰姆矩阵，称 $\det G$ 为该函数向量组的克兰姆行列式。

定理1　函数向量组 $f_i(t)$　$(i\in \underline{n})$线性无关的充要条件为克兰姆行列式 $\det G\neq 0$，即其克兰姆矩阵是正则的。

证　充分性：设 $\det G\neq 0$，则方程组

$$Ga=\theta \qquad a\in R^n$$

只有零解，设

$$\sum_{k=1}^{n} b_k f_k(t)=\theta \qquad b_k\in R, k\in \underline{n} \tag{1}$$

则

$$\int_{t_1}^{t_2} f_i^T(t)\sum_{k=1}^{n} b_k f_k(t)\mathrm{d}t$$

$$=\sum_{k=1}^{n} b_k\int_{t_1}^{t_2} f_i^T(t) f_k(t)\mathrm{d}t$$

$$=\sum_{k=1}^{n} g_{ik}b_k=0 \qquad i\in \underline{n}$$

若记 $b=(b_1,b_2,\cdots,b_n)^T$，即

$$Gb=\theta$$

于是 $b=\theta$，故由(1)式知函数向量组是线性无关的；

必要性：设 $f_i(t)$　$(i\in \underline{n})$线性无关，用反证法，即若 $\det G=0$，则方程组

$$Ga=\theta \qquad a\in R^n \tag{2}$$

有非零解，即当 $a=(a_1,a_2,\cdots,a_n)^T\neq\theta$ 时，

$$\sum_{k=1}^{n} g_{ik}a_k=0 \qquad i\in \underline{n}$$

亦即

$$\sum_{k=1}^{n}\left(\int_{t_1}^{t_2} f_i^T(t) f_k^T(t)\mathrm{d}t\right) a_k = 0 \qquad i \in \underline{n}$$

从而

$$\sum_{i=1}^{n} a_i\left[\sum_{k=1}^{n}\left(\int_{t_1}^{t_2} f_i^T(t) f_k^T(t)\mathrm{d}t\right) a_k\right] = 0$$

上式可改写为

$$\int_{t_1}^{t_2}\left(\sum_{i=1}^{n} a_i f_i^T(t)\right)\left(\sum_{k=1}^{n} a_k f_k(t)\right)\mathrm{d}t$$

$$\int_{t_1}^{t_2}\left\|\sum_{i=1}^{n} a_i f_i(t)\right\|_2^2 \mathrm{d}t = 0$$

于是

$$\sum_{i=1}^{n} a_i f_i(t) = \theta$$

这与函数向量组 $f_i(t) \quad i\in\underline{n}$线性无关的假定矛盾，故

$$\det G \neq 0$$

证毕。

设 $h_i(t) \quad i\in\underline{n}$是在某区间 T 上定义的具有 $n-1$阶连续可导的实函数组，矩阵

$$W(t) \triangleq \begin{bmatrix} h_1(t) & h_1'(t)\cdots h_1^{(n-1)}(t) \\ h_2(t) & h_2'(t)\cdots h_2^{(n-1)}(t) \\ \cdots & \cdots\quad\cdots\quad\cdots\quad\cdots \\ h_n(t) & h_n'(t)\cdots h_n^{n-1}(t) \end{bmatrix}$$

的行列式 $\det W(t)$，称为函数组 $h_i(t) \quad (i\in\underline{n})$的**朗斯基(Wronsky)行列式**，，若 $t\in T$ 时 $\det W(t)\neq 0$，则函数组 $h_i(t) \quad (i\in\underline{n})$是线性无关的。这个结果可推广到函数向量组上去。

定义2 设 $f_i(t)$ $(i\in \underline{n})$是定义在区间$[t_1,t_2]$上的 r 维行函数向量组。记为

$$F(t)=\begin{bmatrix} f_1(t) \\ f_2(t) \\ \cdots\cdots \\ f_n(t) \end{bmatrix}_{n\times r}$$

则称

$$W(t)\triangleq\left[\begin{array}{c:c:c:c} F(t) & F'(t) & \cdots & F^{(n-1)}(t) \end{array}\right]_{n\times nr}$$

为行函数向量组 $f_i(t)$，$i\in \underline{n}$的朗斯基矩阵

定理2 若存在 $t_0\in[t_1,t_2]$使

$$\operatorname{rank}W(t_0)=n$$

则行函数向量组 $f_i(t)$ $(i\in \underline{n})$在$[t_1,t_2]$上线性无关。

证 用反证法。设行函数向量组 $f_i(t)$ $(i\in \underline{n})$在$[t_1,t_2]$上线性相关，则存在 $\theta\neq a\in R^n$，使得

$$a^TF(t)=\theta^T \qquad \forall\ t\in[t_1,t_2]$$

于是

$$a^TF^{(k)}(t)=\theta^T \quad k\in \underline{n-1} \quad \forall\ t\in[t_1,t_2]$$

由此得

$$a^TW(t)=0$$

这与 $\operatorname{rank}W(t_0)=n$ 矛盾。

注意：定理2之逆命题不真，即对于 $t\in[t_1,t_2]$仅$\operatorname{rank}W(t)<n$，行函数向量组 $f_i(t)$ $(i\in \underline{n})$不一定线性相关。其次，若限制 $f_i(t)$的分量均为全平面的解析函数，则可证

$$\operatorname{rank}W(t) = n$$

为 $f_i(t)$ $(i \in \underline{n})$线性无关的充要条件。

习 题 五

一、 设 $A \in F^{n\times n}$，试证$\|A\| < 1$是矩阵序列$\{A^m\}$收敛于O的充分条件。

二、 设$\|A\| < 1$，则

$$\|(I - A)^{-1}\| \leqslant \frac{1}{1 - \|A\|}$$

三、 设 $A \in C^{m\times n}$，则

$$\lim_{k\to+\infty} A^k = I_n$$

的充要条件为 $A = I_n$。

四、 试证尤拉(Euler)公式

$$e^{iAt} = \cos At + i\sin At \qquad i = \sqrt{-1}$$

或

$$\cos At = \frac{1}{2}(e^{iAt} + e^{-iAt})$$

$$\sin At = \frac{1}{2i}(e^{iAt} - e^{-iAt})$$

五、 设

$$A = \begin{pmatrix} 0 & -a \\ a & 0 \end{pmatrix}$$

求 e^A、$\sin A$、$\cos A$。

六、 当 A^{-1}存在时，求证 A^{-1}对 $B_{p\times q}$的导数

$$\frac{\mathrm{d}A^{-1}}{\mathrm{d}B} = -(I_p \otimes A^{-1})\frac{\mathrm{d}A}{\mathrm{d}B}(I_q \otimes A^{-1})$$

七、 证明：

1. 当 $a\in F^m$、$b\in F^n$ 时，有

$$ba^T = a^T \otimes b = b \otimes a^T$$

2. 当 $A=(a_1,a_2,\cdots,a_n)\in F^{m\times n}$时，则

$$A = \sum_{j=1}^{m}(a_j \otimes e_j^T)$$

$$A^T = \sum_{j=1}^{m}(a_j^T \otimes e_j)$$

其中 e_j 为 F^m 中第 j 个坐标为1，其余坐标为0的一组标准正交基，$j\in \underline{m}$。

3. 若$\|x\|_2=\|y\|_2=1$，则

$$\|x \otimes y\|_2 = 1$$

4. 设 $A\in C^{m\times m}$，$B\in C^{n\times n}$，且 $x\in C^m$、$y\in C^n$ 分别是 A 与 B 属于特征值 λ、μ 的特征向量，则 $\lambda\mu$ 是 $A\otimes B$ 的特征值，$x\otimes y$ 是 $A\otimes B$ 对应于特征值 $\lambda\mu$ 的特征向量。

八、

1. 求 Ax 对 x^T 的导数；

2. 求 x^TAx 对 x^T 的导数；

3. 求$(x-a)^TA(x-a)$对 x 的导数；

4. 求$[\sin(x^TAx)]x^T$ 对 x、对 A 的导数；

九、 已知 $B=B^H$，且

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Ax$$

求证

$$\frac{\mathrm{d}}{\mathrm{d}t}(x^TBx) = 2x^TBAx \qquad (A \neq A^H)$$

十、 设 $F(t)=(f_{ij}(t))_{n\times n}=(f_1(t),f_2(t),\cdots,f_n(t))_{n\times n}$ 试证 $f_i(t)$ $(i\in \underline{n})$的克兰姆矩阵为

$$G=\int_{t_1}^{t_2}F^T(t)F(t)\mathrm{d}t$$

十一、 证明函数向量组 $f_i(t)\in R^m$ $(i\in \underline{n})$的克兰姆矩阵是对称的半正定矩阵，若它们是 C^m 上的函数向量，则其克兰姆矩阵是准正定的埃尔米特矩阵。

十二、 设 $f(t)$是$[t_1,t_2]$上的 m 维函数向量，则

$$\left\|\int_{t_1}^{t_2}f(t)\mathrm{d}t\right\|_\infty\leqslant\int_{t_1}^{t_2}\|f(t)\|_\infty\mathrm{d}t$$

式中$\|\cdot\|_\infty$是 R^m 中的范数，上式中的$\|\cdot\|_\infty$可否换为$\|\cdot\|_2$？为什么？

第六章　矩阵函数

上一章，曾就矩阵级数的和函数，给出了一些矩阵函数（例如 e^A、$\sin A$、$\cos A$、$(I-A)^{-1}$等）的表达式。本章，将从更一般的观点给出矩阵函数的定义，并由此导出矩阵函数的更为实用的表达形式。

§6.1　矩阵多项式

正如多项式是函数中最简单的函数一样，矩阵多项式也是矩阵函数中最简单、最重要的一类矩阵函数。

定义1　设 $A\in F^{n\times n}$，

$$f(\lambda)=a_0+a_1\lambda+a_2\lambda^2+\cdots+a_n\lambda^n \tag{1}$$

是 F 上的多项式，则称

$$f(A)=a_0I_n+a_1A+a_2A^2+\cdots+a_nA^n \tag{2}$$

为 A 的**矩阵多项式**。若 (1) 式中 $a_n\neq 0$，则称 n 为**多项式** $f(\lambda)$ **的次数**，简记为 $\deg f(\lambda)=n$。

例1　设 $f(\lambda)=\lambda^2+3\lambda-2$，且

$$A=\begin{pmatrix}1&0\\1&0\end{pmatrix}$$

则

$$f(A)=A^2+3A-2I_2=\begin{pmatrix}1&0\\1&0\end{pmatrix}^2+3\begin{pmatrix}1&0\\1&0\end{pmatrix}-2\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 3 & 0 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 4 & -2 \end{pmatrix}$$

两个矩阵多项式 $f(A)$、$g(A)$ 的和与积分别定义为

$$f(A) + g(A) \triangleq (f + g)(A) \tag{3}$$

$$f(A)g(A) \triangleq (fg)(A) \tag{4}$$

可以看出，对于矩阵多项式，加法运算的交换律、结合律、乘法运算的交换律、结合律以及加法与乘法运算的分配律等，显然都是成立的。

定理1 设 $A \in F^{n\times n}$，$f(\lambda) = \sum_{i=0}^{m} a_i \lambda^i$ 为 F 上的多项式，则

a） $f(T^{-1}AT) = T^{-1}f(A)T$

b） 若 $Ax_i = \lambda_i x_i, \lambda_i \in F, \theta \neq x_i \in F^n$ （$i \in \underline{n}$），则

$$f(A)x_i = f(\lambda_i)x_i \qquad i \in \underline{n}$$

c） 若 A 是单纯矩阵，且其谱分解为

$$A = \sum_{i=1}^{\sigma} \lambda_i E_i$$

则

$$f(A) = \sum_{i=1}^{\sigma} f(\lambda_i) E_i$$

d） 若 $A = \text{diag}(A_1, A_2, \cdots, A_s)$ 为对角块矩阵（其中 A_i，$i \in \underline{s}$ 均为方阵，则

$$f(A) = \text{diag}(f(A_1), f(A_2), \cdots, f(A_s))$$

定理的证明留给读者自己练习。

定义2 设 $A \in F^{n\times n}$，$f(\lambda) = \sum_{i=0}^{m} a_i \lambda^i$ 为 F 上的多项式，若

$$f(A) = O$$

则称 f 为 A 的**化零多项式**。

例2 设 $f(\lambda)=\lambda^2-2\lambda$

$$A=\begin{pmatrix}2 & 2\\ 0 & 0\end{pmatrix}$$

则因

$$f(A)=\begin{pmatrix}2 & 2\\ 0 & 0\end{pmatrix}^2-2\begin{pmatrix}2 & 2\\ 0 & 0\end{pmatrix}=\begin{pmatrix}4 & 4\\ 0 & 0\end{pmatrix}-\begin{pmatrix}4 & 4\\ 0 & 0\end{pmatrix}=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$$

故 $f(\lambda)=\lambda^2-2\lambda$ 是 $A=\begin{pmatrix}2 & 2\\ 0 & 0\end{pmatrix}$ 的化零多项式。容易验证 $g(\lambda)=\lambda^3-4\lambda$ 也是 $A=\begin{pmatrix}2 & 2\\ 0 & 0\end{pmatrix}$ 的化零多项式。由此可见，一个矩阵的化零多项式并不唯一。

化零多项式具有下面的性质。

定理2 a) 任意 n 阶方阵 A 均有化零多项式；

b) A 有次数最低的化零多项式。

证 a) 由于 $F^{n\times n}$ 是 n^2 维线性空间，故

$$I_n、A、A^2、\cdots、A^{n^2}$$

必线性相关，即存在 n^2+1 个不全为零的数 $a_0、a_1、a_2、\cdots、a_{n^2}$，使得

$$a_0I_n+a_1A+a_2A^2+\cdots+a_{n^2}A^{n^2}=O$$

即 $f(\lambda)=a_0+a_1\lambda+a_2\lambda^2+\cdots+a_{n^2}\lambda^{n^2}$ 为 A 的化零多项式。

b) 由正整数集均有最小正整数，即知结论正确。

定义3 矩阵 $A\in F^{n\times n}$ 的次数最低的化零多项式，称为矩阵 A 的**最小多项式**，记为 $\psi_m(\lambda)$。

矩阵 A 的最小多项式具有以下性质。

定理 3　矩阵 $A \in F^{n\times n}$，则

a)　A 的任意化零多项式均能被 A 的最小多项式整除；

b)　A 的首1（多项式的最高次项系数为1）最小多项式唯一；

c)　相似矩阵的最小多项式相同；

d)　设 $\psi_m(\lambda)$ 为 $A \in C^{n\times n}$ 的首 1 最小多项式，$A = TJT^{-1}$ 为 A 的约当标准形分解，

$$J = \mathrm{diag}(J_1, J_2, \cdots, J_\sigma)$$

$$J_i = \mathrm{diag}(J_{i1}, J_{i2}, \cdots, J_{i\alpha_i}) \quad i \in \underline{\sigma}$$

$J_i \in C^{m_i\times m_i}$，$J_{ik} \in C^{n_{ik}\times n_{ik}}$，$\sigma$ 为 A 相异特征值的个数，m_i、α_i($i \in \underline{\sigma}$) 分别为 A 的特征值 λ_i 的代数重复度与几何重复度，n_{ik} 为 A 的初等因子 $\lambda - \lambda_i$ 的幂，若 $d_i = \max\limits_k n_{ik}$，则

$$\psi_m(\lambda) = (\lambda - \lambda_1)^{d_1}(\lambda - \lambda_2)^{d_2}\cdots(\lambda - \lambda_\sigma)^{d_\sigma}$$

证　a) 设 $f(\lambda)$ 为 A 的化零多项式，$\psi_m(\lambda)$ 为 A 的最小多项式，由最小多项式的定义知

$$\deg f(\lambda) \geqslant \deg \psi_m(\lambda)$$

故由多项式的带余数除法知，存在多项式 $q(\lambda)$、$r(\lambda)$，使得

$$f(\lambda) = \psi_m(\lambda)q(\lambda) + r(\lambda) \tag{5}$$

其中或者 $r(\lambda) = 0$，或者 $r(\lambda) \neq 0$. 但必须有

$$\deg r(\lambda) < \deg \psi_m(\lambda)$$

若 $r(\lambda) \neq 0$，则由 $f(A) = O$、$\psi_m(A) = O$ 及 (5) 式有

$$r(A) = O$$

这与 $\psi_m(\lambda)$ 是 A 的最小多项式矛盾，所以 $r(\lambda) = 0$，即

$$\psi_m(\lambda) \mid f(\lambda)$$

b)　设 $\psi_m(\lambda)$ 与 $\varphi(\lambda)$ 均为 A 的首1最小多项式，则由

$$\varphi(\lambda)=\psi_m(\lambda)q(\lambda),\psi_m(\lambda)=\varphi(\lambda)p(\lambda)$$

得

$$\varphi(\lambda)=\varphi(\lambda)p(\lambda)q(\lambda)$$

由于$\varphi(\lambda)$是首1的，故$p(\lambda)q(\lambda)=1$，所以$p(\lambda)=q(\lambda)=1$，即$\varphi(\lambda)=\psi_m(\lambda)$；

c)　设$B=T^{-1}AT$，由于

$$\psi_m(B)=T^{-1}\psi_m(A)T$$

故A与B有相同的最小多项式（$\psi_m(A)=O$与$\psi_m(B)=O$等价）；

d)　设A的最小多项式为

$$\psi_m(\lambda)=(\lambda-\mu_1)^{d_1}(\lambda-\mu_2)^{d_2}\cdots(\lambda-\mu_s)^{d_s}$$

$A=TJT^{-1}$为矩阵A的约当标准形分解，于是$\psi_m(A)=O$等价于$\psi_m(J)=O$，若记A的约当块为

$$J_j=\begin{bmatrix} J_{j1} & & & \\ & J_{j2} & & \\ & & \ddots & \\ & & & J_{j\alpha_j} \end{bmatrix}_{m_j\times m_j} \qquad j\in\underline{\sigma}$$

显然，$\psi_m(J)=O$等价于$\psi_m(J_j)=O,j\in\underline{\sigma}$，而$\psi_m(J_j)=O$又等价于$\psi_m(J_{jk})=O,j\in\underline{\sigma},k\in\underline{\alpha_j}$，这只有$\mu_1$、$\mu_2$、$\cdots$、$\mu_s$全为$A$的特征值，且$d_j\geqslant\max\limits_k n_{jk}$（其中$n_{jk}$为约当子块$J_{jk}$的阶数）才可能。由于最小多项式是次数最低的化零多项式，故取$d_j=\max\limits_k n_{jk}$，于是A的最小多项式为

$$\psi_m(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_\sigma)^{d_\sigma} \tag{6}$$

今后凡是提到最小多项式，都是首1的。

定理 4 设 $A\in C^{n\times n}$，则

a）凯莱-哈米尔顿（Cayley-Hamilton）定理

A 的特征多项式是 A 的化零多项式；

b）A 的所有特征值的几何重复度均为 1 的充要条件，为 A 的特征多项式与最小多项式一致；

c）A 是单纯矩阵的充要条件为 $d_j=1$，$j\in \underline{\sigma}$；

d）若 $d(\lambda)$ 为 $\operatorname{adj}(\lambda I-A)$ 全部元素的最大公因式，则

$$\psi(\lambda)=\psi_m(\lambda)d(\lambda)$$

其中 $\psi(\lambda)$ 为 A 的特征多项式；$\psi_m(\lambda)$ 为 A 的最小多项式。

证 *a*）由定理 3 的 *d*）中所得最小多项式的构造即知；

b)、*c*）显然；

d）设 $\operatorname{adj}(\lambda I-A)=C(\lambda)d(\lambda)$，代入

$$(\lambda I-A)\operatorname{adj}(\lambda I-A)=\psi(\lambda)I_n$$

得

$$(\lambda I-A)C(\lambda)d(\lambda)=\psi(\lambda)I_n$$

显然 $d(\lambda)\,|\,\psi(\lambda)$，为此，令 $\psi(\lambda)=d(\lambda)\varphi(\lambda)$，代入上式得

$$(\lambda I-A)C(\lambda)=\varphi(\lambda)I_n \tag{7}$$

由此可知 $\varphi(\lambda)$ 是 A 的化零多项式，因此，可令

$$\varphi(\lambda)=\psi_m(\lambda)q(\lambda)$$

由于 $\psi_m(A)=0$，故

$$\psi_m(\lambda)I_n=(\lambda I-A)\tilde{C}(\lambda)$$

于是用 $q(\lambda)$ 乘上式两端得

$$\varphi(\lambda)I_n=(\lambda I-A)\tilde{C}(\lambda)q(\lambda) \tag{8}$$

比较（7）与（8），并由除法的唯一性得

$$C(\lambda)=\tilde{C}(\lambda)q(\lambda)$$

由此可见 $q(\lambda)$ 是 $C(\lambda)$ 全部元素的公因式，但 $C(\lambda)$ 的最大公

因式只能是 1，即 $\varphi(\lambda) = \psi_m(\lambda)$，亦即

$$\psi(\lambda) = \psi_m(\lambda) d(\lambda)$$

证毕。

定理 5 设 $f(\lambda)$、$g(\lambda)$ 为 λ 的两个多项式，$A \in C^{n\times n}$，则 $f(A) = g(A)$ 的充要条件为

$$f^{(l)}(\lambda_j) = g^{(l)}(\lambda_j), j \in \underline{\sigma}, l = 0, 1, \cdots, d_j - 1 \qquad (9)$$

其中 $f^{(l)}(\lambda_j) = \left.\frac{d^l f(\lambda)}{d\lambda^l}\right|_{\lambda=\lambda_j}$，$d_j = \max\limits_k n_{jk}$ $(j \in \underline{\sigma})$ 为 A 的最小多项式的各个一次因式的幂。

证 $f(A) = g(A)$ 等价于 $f(A) - g(A) = O$，它又等价于 $f(\lambda) - g(\lambda) = \psi_m(\lambda) q(\lambda)$，显然这又与 (9) 式等价。证毕。

推论 $f(A) = O$ 的充要条件为

$$f^{(l)}(\lambda_j) = 0 \quad j \in \underline{\sigma}, \quad l = 0, 1, \cdots, d_j - 1$$

定义 4 设 $A \in C^{n\times n}$，$f(\lambda)$、$g(\lambda)$ 为域 C 上的两个多项式，$\lambda_i (i \in \underline{\sigma})$ 为 A 的相异特征值，$d_i = \max\limits_k n_{ik}$，$i \in \underline{\sigma}$，若

$$f^{(l)}(\lambda_i) = g^{(l)}(\lambda_i) \quad i \in \underline{\sigma}, \quad l = 0, 1, \cdots, d_j - 1$$

则称 $f(\lambda)$ 与 $g(\lambda)$ 在 A 的谱上一致。

定理 6 设 $A \in C^{n\times n}$，$f(\lambda)$ 为域 C 上的多项式，则

$$f(A) = \sum_{i=1}^{\sigma} \sum_{l=0}^{d_i - 1} \left.\left(\frac{\mathrm{d}^l f}{\mathrm{d}\lambda^l}\right)\right|_{\lambda=\lambda_i} E_i^l$$

其中

$$E_i^l = T H_i^l T^{-1}$$

$$H_i^0 = \begin{bmatrix} O & & & & & \\ & \ddots & & & & \\ & & O & & & \\ & & & I_{m_i} & & \\ & & & & O & \\ & & & & & \ddots & \\ & & & & & & O \end{bmatrix} \leftarrow \text{第 } i \text{ 个子块}$$

$$H_i^1 = (\hat{J}_i - \lambda_i H_i^0), H_i^l = \frac{1}{l!}(H_i^1)^l (l \geqslant 2)$$

$$\hat{J}_i = \begin{pmatrix} O & & & & \\ & \ddots & & & \\ & & O & & \\ & & & J_i & & \\ & & & & O & \\ & & & & & \ddots \\ & & & & & & O \end{pmatrix} \leftarrow \text{第 } i \text{ 个子块}$$

$A = TJT^{-1}$ 为 A 的约当标准形分解。

证　设 $A = TJT^{-1}, J = \sum_{i=1}^{\sigma} \hat{J}_i$ 及

$$g(\lambda) = f(\lambda_i) + f'(\lambda_i)(\lambda - \lambda_i) + \frac{f''(\lambda_i)}{2!}(\lambda - \lambda_i)^2$$

$$+ \cdots + \frac{f^{(d_i-1)}(\lambda_i)}{(d_i - 1)!}(\lambda - \lambda_i)^{d_i-1}$$

于是 $f(\lambda)$ 与 $g(\lambda)$ 在 A 的谱上一致，从而

$$f(J_i) = g(J_i) = f(\lambda_i)I_{m_i} + f'(\lambda_i)(J_i - \lambda_i I_{m_i})$$

$$+ \frac{f''(\lambda_i)}{2!}(J_i - \lambda_i I_{m_i})^2 + \cdots + \frac{f^{(d_i-1)}(\lambda_i)}{(d_i - 1)!}(J_i - \lambda_i I_{m_i})^{d_i-1}$$

于是

$$f(\hat{J}_i) = f(\lambda_i)H_i^0 + f'(\lambda_i)H_i^1 + \cdots + f^{(d_i-1)}(\lambda_i)H_i^{d_i-1}$$

故

$$f(A) = Tf(J)T^{-1} = T\Big(\sum_{i=1}^{\sigma} f(\hat{J}_i)\Big)T^{-1}$$

$$= \sum_{i=1}^{\sigma} T\Big(\sum_{l=0}^{d_i-1} f^{(l)}(\lambda_i)H_i^l\Big)T^{-1} = \sum_{i=1}^{\sigma}\sum_{l=0}^{d_i-1}\left(\frac{\mathrm{d}^l f}{\mathrm{d}\lambda^l}\right)\Bigg|_{\lambda=\lambda_i} E_i^l$$

证毕。

例 3　设

$$A=\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\end{bmatrix}$$

试用凯莱-哈米尔顿定理．求 $A^7-A^3+2I_4$。

解 $\psi(\lambda)=\det(\lambda I-A)=\lambda^4-1$，于是由凯莱-哈米尔顿定理，知 $A^4-I_4=O$，故

$$A^7-A^3+2I_4=A^3(A^4-I_4)+2I_4=2I_4$$

例4 设

$$A=\begin{bmatrix}2&1&1\\1&2&1\\1&1&2\end{bmatrix},$$

用凯莱-哈米尔顿定理求 A^{-1}。

解 $\psi(\lambda)=\det(\lambda I-A)=\lambda^3-6\lambda^2+9\lambda-4$，由凯莱-哈米尔顿定理知 $A^3-6A^2+9A=4I_3$。于是

$$A^{-1}=\frac{1}{4}(A^2-6A+9I_3)=\frac{1}{4}\begin{bmatrix}3&-1&-1\\-1&3&-1\\-1&-1&3\end{bmatrix}$$

例5 设 $A=TJT^{-1}$

$$J=\left\{\begin{array}{ccc:c:c:cc}\lambda_1&1&&&&&\\&\lambda_1&1&&&&\\&&\lambda_1&&&&\\ \hdashline &&&\lambda_1&&&\\ \hdashline &&&&\lambda_1&&\\ \hdashline &&&&&\lambda_2&1\\&&&&&&\lambda_2\end{array}\right\}$$

$f(\lambda)=(\lambda-1)^2$，求 $f(A)$。

解 由于知 $m_1=5$，$m_2=2$，$d_1=3$，$d_2=2$，于是由定理6知

$$H_1^0=\left(\begin{array}{ccccc:cc} 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & \multicolumn{2}{c}{O} \\ & & & 1 & & & \\ & & & & 1 & & \\ \hdashline \multicolumn{5}{c:}{O} & \multicolumn{2}{c}{O} \end{array}\right)_{7\times 7}$$

$$H_2^0=\left(\begin{array}{c:cc} O & \multicolumn{2}{c}{O} \\ \hdashline O & 1 & \\ & & 1 \end{array}\right)_{7\times 7}$$

$$H_1^1=\left(\begin{array}{ccc:c:c} 0 & 1 & 0 & & \\ 0 & 0 & 1 & O & \\ 0 & 0 & 0 & & O \\ \hdashline \multicolumn{3}{c:}{O} & O & \\ \hdashline \multicolumn{4}{c:}{O} & O \end{array}\right)_{7\times 7}$$

$$H_1^2=\left(\begin{array}{ccc:c} 0 & 0 & 1 & O \\ 0 & 0 & 0 & \\ \hdashline \multicolumn{3}{c:}{O} & O \end{array}\right)_{7\times 7}$$

$$H_2^1=\left(\begin{array}{c:cc} O & \multicolumn{2}{c}{O} \\ \hdashline O & 0 & 1 \\ & 0 & 0 \end{array}\right)_{7\times 7}$$

$$f'(\lambda)=2(\lambda-1),f''(\lambda)=2$$

于是

$$
\begin{aligned}
f(A) &= \sum_{i=1}^{2}\sum_{l=0}^{d_i-1} f^{(l)}(\lambda_i)\, E_i^l \\
&= (\lambda_1 - 1)^2 E_1^0 + 2(\lambda_1 - 1)E_1^1 + 2E_1^2 \\
&\quad + (\lambda_2 - 1)^2 E_2^0 + 2(\lambda_2 - 1)E_2^1 \\
&= T[(\lambda_1 - 1)^2 H_1^0 + 2(\lambda_1 - 1)H_1^1 + 2H_1^2 \\
&\quad + (\lambda_2 - 1)^2 H_2^0 + 2(\lambda_2 - 1)H_2^1]T^{-1}
\end{aligned}
$$

$$
= T\begin{bmatrix}
(\lambda_1-1)^2 & 2(\lambda_1-1) & 1 & & & & \\
 & (\lambda_1-1)^2 & 2(\lambda_1-1) & & & & \\
 & & (\lambda_1-1)^2 & & & & \\
 & & & (\lambda_1-1)^2 & & & \\
 & & & & (\lambda_1-1)^2 & & \\
 & & & & & (\lambda_1-1)^2 & 2(\lambda_1-1) \\
 & & & & & & (\lambda_1-1)^2
\end{bmatrix} T^{-1}
$$

§ 6.2 矩阵函数的定义与性质

首先，利用矩阵多项式给出矩阵函数的定义，然后，再讨论矩阵函数的性质。

定义 设 $f(z)$为域 C 中在包含 $A\in C^{n\times n}$的特征值的开域上的解析函数，$g(\lambda)$为域 C 上的多项式，且 $g(\lambda)$与 $f(\lambda)$在 A 的谱上一致，则规定

$$f(A) \triangleq g(A)$$

由上节可知，对于给定的解析函数 $f(z)$，一定存在与它在 A 的谱上一致的多项式 $g(\lambda)$，如果 $g_1(\lambda)$、$g_2(\lambda)$是在 A 的谱上与 $f(z)$一致的任意两个多项式，显然 $g_1(A)=g_2(A)$。基于这一点，可用 A 的特征多项式，代替 A 的最小多项式来定义 $f(A)$，这样，虽然项数可能多算一些，但省去了求最小多项式。

例 1 设

$$A=\begin{bmatrix}\lambda_0 & 1 & & & \\ & \lambda_0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_0\end{bmatrix}_{m\times m}$$

$f(z)$为域 C 上的解析函数，求 $f(A)$。

解 由于 A 只有一个 m 重特征值 λ_0，故其几何重复度显然为 1。由上节定理 4 之 b)知，A 的特征多项式与最小多项式一致，即

$$\psi_m(\lambda)=\psi(\lambda)=(\lambda-\lambda_0)^m$$

令

$$g(\lambda)=f(\lambda_0)+f'(\lambda_0)(\lambda-\lambda_0)+\cdots+\frac{f^{(m-1)}(\lambda_0)}{(m-1)!}(\lambda-\lambda_0)^{m-1}$$

于是 $g(\lambda)$是与 $f(z)$在 A 的谱上一致的多项式，故由矩阵函数定义，有

$$f(A)=g(A)=\begin{bmatrix}f(\lambda_0) & f'(\lambda_0) & \frac{1}{2!}f''(\lambda_0) & \cdots & \frac{1}{(m-1)!}f^{(m-1)}(\lambda_0) \\ & f(\lambda_0) & f'(\lambda_0) & \cdots & \frac{1}{(m-2)!}f^{(m-2)}(\lambda_0) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda_0) \\ & & & & f(\lambda_0)\end{bmatrix}$$

矩阵函数具有以下性质。

定理 1 设 $A\in C^{n\times n}$，$f_1(z)$、$f_2(z)$是域 C 上的解析函数。则

a) $Af_1(A)=f_1(A)A$

b) $(f_1+f_2)(A)=f_1(A)+f_2(A)$

c) $(f_1f_2)(A)=f_1(A)f_2(A)$

d) 若 $f_1(z)$与 $f_2(z)$在 A 的谱上一致，则

$$f_1(A) = f_2(A)$$

e) 若 $B\in C^{n\times n}$，且 $B=TAT^{-1}$，则

$$f_1(B) = Tf_1(A)T^{-1}$$

f) 设

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & & & \\ & \boldsymbol{A}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{A}_s \end{pmatrix} = \boldsymbol{A}_1 \oplus \boldsymbol{A}_2 \oplus \cdots \oplus \boldsymbol{A}_s$$

$$= \mathrm{diag}(\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_s)$$

则

$$\begin{aligned} f_1(A) &= f_1(A_1) \oplus f_1(A_2) \oplus \cdots \oplus f_1(A_s) \\ &= \mathrm{diag}(f_1(A_1), f_1(A_2), \cdots, f_1(A_s)) \end{aligned}$$

以上只要注意到矩阵多项式具有相应的性质即得。

定理 2 设 $g(x_1, x_2)$是变量 x_1、x_2 的多项式，$f_1(z)$、$f_2(z)$在 A 的谱上有

$$g(f_1(\lambda), f_2(\lambda)) = 0$$

则

$$g(f_1(A), f_2(A)) = O$$

证 设与 $f_1(z)$、$f_2(z)$在 A 的谱上一致的多项式为 $g_1(\lambda)$与 $g_2(\lambda)$，且

$$g(g_1(\lambda), g_2(\lambda)) = q(\lambda)$$

则在 A 的谱上有

$$q(\lambda) = g(g_1(\lambda), g_2(\lambda)) = g(f_1(\lambda), f_2(\lambda)) = 0 \quad \lambda \in \lambda(A)$$

于是，由上节定理 5 的推论知

$$q(A) = O$$

即

$$g(g_1(A), g_2(A)) = g(f_1(A), f_2(A)) = O$$

例 2 试证对任意的 $A \in C^{n \times n}$，有

a) $\cos^2 A + \sin^2 A = I_n$

b) $e^A \cdot e^{-A} = I_n$

c) $e^{iA} = \cos A + i\sin A \qquad i = \sqrt{-1}$

证 a) 设 $g(x_1, x_2) = x_1^2 + x_2^2 - 1$，$f_1(z) = \cos z$，$f_2(z) = \sin z$，则由

$$\cos^2 z + \sin^2 z = 1, (\cos^2 z + \sin^2 z)' = 0$$

有

$$g(f_1(z), f_2(z)) = 0 \qquad \forall z \in C$$

$$g^{(l)}(f_1(z), f_2(z)) = 0 \qquad \forall z \in C$$

由定理 2 有

$$g(f_1(A), f_2(A)) = O$$

即

$$\cos^2 A + \sin^2 A = I_n$$

b) 设 $g(x_1, x_2) = x_1 x_2 - 1$，$f_1(z) = e^z$、$f_2(z) = e^{-z}$则由 $e^z, e^{-z} = 1$ 有

$$g(f_1(z), f_2(z)) = 0$$

故

$$g(f_1(A), f_2(A)) = O$$

即

$$e^A e^{-A} = I_n$$

c) 设 $g(x_1,x_2)=x_1-x_2$，$f_1(z)=\cos z+i\sin z$，$f_2(z)=e^{iz}$，则

$$g(f_1(z),f_2(z))=0$$

故

$$g(f_1(A),f_2(A))=O$$

即

$$e^{iA}=\cos A+i\sin A$$

§6.3 $f(A)$用 Jordan 标准形表示（标准形Ⅰ）

从本节开始，介绍矩阵函数 $f(A)$的几种常用的表示方法，先介绍用 A 的约当标准形 J 表示 $f(A)$，§6.1 曾介绍过用约当标准形表示矩阵多项式，故此处只是将该结果推广到一般的矩阵函数上去。

定理 设 $A\in C^{n\times n}$，$A=TJT^{-1}$为 A 的约当标准形分解，$f(z)$为域 C 上的解析函数，则

$$f(A)=Tf(J)T^{-1} \tag{1}$$

其中

$$f(J)=\mathrm{diag}(f(J_1),\ f(J_2),\cdots,f(J_\sigma))$$

$$f(J_i)=\mathrm{diag}(f(J_{i1}),\ f(J_{i2}),\cdots,f(J_{i\,\sigma_i}))\ i\in\underline{\sigma}$$

$$f(J_{ik})=\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{f^{(n_{ik}-1)}(\lambda_i)}{(n_{ik}-1)!} \\ & f(\lambda_i) & f'(\lambda_i) & \cdots\cdots & \frac{f^{(n_{ik}-2)}(\lambda_i)}{(n_{ik}-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda_i) \\ & & & & f(\lambda_i) \end{bmatrix}$$

定理成立是显然的。我们称 $f(A)$的这种表示式为$f(A)$的标准形Ⅰ。

特别，若 A 为单纯矩阵，则 J 为对角线矩阵，显然此时 $f(J)$ 也是对角线矩阵，所以相应的 $f(A)$ 的表示式也更为简单。

例 1 设 $A\in C^{n\times n}$，求矩阵函数 e^{At} 的标准形Ⅰ。

解 因为 e^{xt} 是域 C 上的解析函数，由上述定理有

$$e^{At}=Te^{Jt}T^{-1}$$

其中

$$e^{Jt}=\begin{bmatrix} e^{J_1t} & & & \\ & e^{J_2t} & & \\ & & \ddots & \\ & & & e^{J_\sigma t}\end{bmatrix}_{n\times n}$$

$$e^{J_it}=\begin{bmatrix} e^{J_{i1}t} & & & \\ & e^{J_{i2}t} & & \\ & & \ddots & \\ & & & e^{J_{i\sigma_i}t}\end{bmatrix}_{m_i\times m_i}\qquad i\in\underline{\sigma}$$

$$e^{J_{ik}t}=\begin{bmatrix} e^{\lambda_it} & te^{\lambda_it} & \frac{t^2}{2!}e^{\lambda_it} & \cdots\cdots & \frac{t^{n_{ik}-1}}{(n_{ik}-1)!}e^{\lambda_it} \\ & e^{\lambda_it} & te^{\lambda_it} & \cdots\cdots & \frac{t^{n_{ik}-2}}{(n_{ik}-2)!}e^{\lambda_it} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & te^{\lambda_it} \\ & & & & e^{\lambda_it}\end{bmatrix}$$

$$i\in\underline{\sigma},k\in\underline{\sigma_i}$$

例 2 设

$$A=\begin{bmatrix}3 & 1 & -1\\1 & 2 & -1\\2 & 1 & 0\end{bmatrix}$$

求 e^{At} 的标准形 I 。

解 由于

$$\lambda I-A=\begin{bmatrix}\lambda-3 & -1 & 1\\-1 & \lambda-2 & 1\\-2 & -1 & \lambda\end{bmatrix}\to\begin{bmatrix}1 & 0 & 0\\0 & \lambda-1 & 0\\0 & 0 & (\lambda-2)^2\end{bmatrix}$$

所以 $\lambda_1=1$，$\lambda_2=\lambda_3=2$，$d_1=1$，$d_2=2$，于是 A 的约当标准形为

$$J=\begin{bmatrix}1 & & \\ & 2 & 1\\ & & 2\end{bmatrix}$$

由于

$$I-A=\begin{bmatrix}-2 & -1 & 1\\-1 & -1 & 1\\-2 & -1 & 1\end{bmatrix}$$

求出对应 $\lambda_1=1$ 的特征向量 $t_1=(0, 1, 1)^T$，又由

$$2I-A=\begin{bmatrix}-1 & -1 & 1\\-1 & 0 & 1\\-2 & -1 & 2\end{bmatrix}$$

求出对应 $\lambda_2=\lambda_3=2$ 的特征向量 $t_2=(1, 0, 1)^T$ 及其广义特征向量 $t_3=(1, 1, 1)^T$，于是

$$T=(t_1, t_2, t_3)=\begin{bmatrix}0 & 1 & 1\\1 & 0 & 1\\1 & 1 & 1\end{bmatrix}$$

$$T^{-1}=\begin{bmatrix}-1 & 1 & 1\\ 0 & -1 & 1\\ 1 & 1 & -1\end{bmatrix}$$

故得

$$e^{At}=Te^{Jt}T^{-1}=\begin{bmatrix}0 & 1 & 1\\ 1 & 0 & 1\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix}e^{t} & & \\ & e^{2t} & te^{2t}\\ & & e^{2t}\end{bmatrix}\begin{bmatrix}-1 & 1 & 1\\ 0 & -1 & 1\\ 1 & 1 & -1\end{bmatrix}$$

$$=\begin{bmatrix}(t+1)e^{2t} & te^{2t} & (t+2)e^{2t}\\ -e^{t}+e^{2t} & e^{t}+e^{2t} & e^{t}+e^{2t}\\ -e^{t}+(t+1)e^{2t} & e^{t}+te^{2t} & e^{t}+(t+2)e^{2t}\end{bmatrix}$$

例3 设

$$A=\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}$$

求 e^{At}的标准形 I 。

解 $\det(\lambda I-A)=(\lambda-1)(\lambda-3)$，故

$$J=\begin{bmatrix}1 & 0\\ 0 & 3\end{bmatrix}$$

且易求得 A 对应于 $\lambda_1=1$，$\lambda_2=3$ 的特征向量分别为

$$t_1=(1,\ 1)^T,\ t_2=(-1,\ 1)^T$$

于是

$$T=\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix},T^{-1}=\frac{1}{2}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}$$

故代入(1)得

$$e^{At}=Te^{Jt}T^{-1}$$

$$=\begin{bmatrix}1 & -1\\1 & 1\end{bmatrix}\begin{bmatrix}e^t & 0\\0 & e^{3t}\end{bmatrix}\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\-\frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

$$=\frac{1}{2}\begin{bmatrix}e^t+e^{3t} & e^t-e^{3t}\\e^t-e^{3t} & e^t+e^{3t}\end{bmatrix}$$

§6.4 $f(A)$用拉格朗日-西勒维斯特(Lagrange-Sylvester)内插多项式表示(标准形Ⅰ)

由于矩阵函数是用矩阵多项式来定义的，所以寻求矩阵函数的表示式，便可转化为求与它在矩阵 A 的谱上一致的矩阵多项式，这样的矩阵多项式不唯一。本节将给出一个求比 A 的最小多项式的次数低的与 $f(z)$在 A 的谱上一致的多项式 $p(\lambda)$来表示 $f(A)$的具体方法。这里的 $p(\lambda)$将称为**拉格朗日-西勒维斯特内插多项式**。

为书写方便起见，用

$$f(\Lambda(A)) = g(\Lambda(A))$$

表示 $f(z)$与 $g(z)$在 A 的谱上一致，即

$$f^{(l)}(\lambda_i)=g^{(l)}(\lambda_i) \quad i\in\underline{\sigma}, l=0,1,2,\cdots,d_i-1$$

其中 $\Lambda(A)$表示 A 的谱集，$\lambda_i\in\Lambda(A) \quad (i\in\underline{\sigma})$，若

$$f^{(l)}(\lambda_i)=0 \quad i\in\underline{\sigma}, l=0,1,2,\cdots,d_i-1$$

记为

$$f(\Lambda(A))=O$$

定理1 设 $A\in C^{n\times n}$，$p(\lambda)$为多项式，则

$$f(\Lambda(A))=O$$

的充要条件为 $\psi_m(\lambda)\mid p(\lambda)$，其中 $\psi_m(\lambda)$为 A 的最小多项式。

证 由

$$f^{(l)}(\lambda_i)=0 \quad l=0,1,2,\cdots,d_i-1(i\in \underline{\sigma})$$

等价于 λ_i 是 $p(\lambda)$ 的 $d_i'(d_i'\geqslant d_i)$ 重根，故等价于

$$(\lambda-\lambda_i)^{d_i}|p(\lambda) \qquad i\in \underline{\sigma}$$

也等价于

$$\psi_m(\lambda)=(d-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_\sigma)^{d_\sigma}|p(\lambda)$$

定理 2 设 $A\in C^{n\times n}$，$f(z)$ 为在 $\Lambda(A)$ 上给定的多项式，且 $f(\Lambda(A))\neq O$，若有多项式 $p_1(\lambda)$，使得

$$f(\Lambda(A)) = p_1(\Lambda(A))$$

则存在多项式 $p(\lambda)$，使得

$$f(\Lambda(A)) = p(\Lambda(A)) \quad \deg p < \deg\psi_m = m$$

其中 $\psi_m(\lambda)$ 为 A 的最小多项式。

证 不妨设 $\deg p_1(\lambda)\geqslant m$，则有

$$p_1(\lambda) = \psi_m(\lambda)q(\lambda) + r(\lambda)$$

其中 $\deg r(\lambda)<m$，由于 $f(\Lambda(A))\neq O$，有 $r(\Lambda(A))\neq O$，且

$$f(\Lambda(A)) = p_1(\Lambda(A)) = r(\Lambda(A)) \quad (\psi_m(\Lambda(A)) = O)$$

于是取 $p(\lambda)=r(\lambda)$ 即可。

定理 1 表明，在 A 的谱上与 O 一致的多项式一定是 A 的化零多项式，而且化零多项式定义的矩阵函数，无疑是零函数；定理 2 表明当 $f(\Lambda(A))\neq O$ 时，一定存在次数比 A 的最小多项式次数低的多项式 $p(\lambda)$，使得

$$f(\Lambda(A)) = p(\Lambda(A))$$

下面给出求 $p(\lambda)$ 的具体方法。

定理 3 设 $A\in C^{n\times n}$ 的最小多项式为

$$\psi_m(\lambda) = (\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_\sigma)^{d_\sigma}$$

$$\deg\psi_m(\lambda) = m = \sum_{i=1}^{\sigma} d_i$$

则多项式

$$p(\lambda) = \sum_{i=1}^{\sigma}\Big(\sum_{l=1}^{d_i}\alpha_{il}(\lambda-\lambda_i)^{l-1}\Big)\varphi_i(\lambda)$$

其中 $\varphi_i(\lambda) = \dfrac{\psi_m(\lambda)}{(\lambda-\lambda_i)^{d_i}} = \prod_{\substack{j=1\\j\neq i}}^{\sigma}(\lambda-\lambda_j)^{d_j} \qquad i\in\underline{\sigma}$

$$\deg p(\lambda) < m$$

且

$$f(\Lambda(A)) = p(\Lambda(A))$$

证　设 $\deg p(\lambda)<m$，则 $\dfrac{p(\lambda)}{\psi_m(\lambda)}$ 可展成部分分式之和，即

$$\frac{p(\lambda)}{\psi_m(\lambda)} = \sum_{i=1}^{\sigma}\left[\frac{\alpha_{i1}}{(\lambda-\lambda_i)^{d_i}} + \frac{\alpha_{i2}}{(\lambda-\lambda_i)^{d_i-1}} + \cdots + \frac{\alpha_{id_i}}{\lambda-\lambda_i}\right] \quad (1)$$

由留数计算公式知

$$\alpha_{il} = \frac{1}{(l-1)!}\frac{\mathrm{d}^{l-1}}{\mathrm{d}\lambda^{l-1}}(\lambda-\lambda_i)^{d_i}\frac{p(\lambda)}{\psi_m(\lambda)}\bigg|_{\lambda=\lambda_i} \quad l\in\underline{d_i}, i\in\underline{\sigma}$$

此式表明系数 α_{il} 只与 $P(\lambda)$ 在谱上的值有关。又因

$$f(\Lambda(A)) = p(\Lambda(A))$$

故上式又可写为

$$\alpha_{il} = \frac{1}{(l-1)!}\frac{\mathrm{d}^{l-1}}{\mathrm{d}\lambda^{l-1}}(\lambda-\lambda_i)^{d_i}\frac{f(\lambda)}{\psi_m(\lambda)}\bigg|_{\lambda=\lambda_i} l\in\underline{d_i}, i\in\underline{\sigma} \quad (2)$$

将所求出的系数 α_{il} 代入(1)式可得

$$p(\lambda) = \sum_{i=1}^{\sigma}\Big(\sum_{l=1}^{d_i}\alpha_{il}(\lambda-\lambda_i)^{l-1}\Big)\varphi_i(\lambda)$$

这里 $f(\Lambda(A))=p(\Lambda(A))$ 是容易验证的。由此可定义

$$f(A) = p(A) = \sum_{i=1}^{\sigma}\Big(\sum_{l=1}^{d_i}\alpha_{il}(A-\lambda_i I)^{l-1}\Big)\varphi_i(A) \quad (3)$$

此处系数 α_{il} 由(2)式给出；(3)式称为矩阵函数 $f(A)$ 的拉格

朗日-西勒维斯特内插多项式。$f(A)$的这种表示式称为 $f(A)$ 的标准形Ⅱ。

作为特例，当 $d_i=1$，$i\in\underline{\sigma}$ 时，则

$$\alpha_i=\frac{f(\lambda_i)}{\varphi_i(\lambda_i)} \qquad i\in\underline{\sigma}$$

$$\varphi_i(\lambda)=\prod_{\substack{j=1\\j\neq i}}^{\sigma}(\lambda-\lambda_j)$$

$$f(A)=\sum_{i=1}^{\sigma}\frac{f(\lambda_i)}{\varphi_i(\lambda_i)}\varphi_i(A) \tag{4}$$

这也包括 A 有 n 个相异特征值的特殊情形，这时，只须在(4)式中令 $\sigma=n$ 即可。

例 设

$$A=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ -6 & -11 & -6\end{bmatrix}$$

求转移矩阵 e^{At}。

$a)$ 用标准形Ⅰ；

$b)$ 用标准形Ⅱ。

解 $a)$ $\det(\lambda I-A)=(\lambda+1)(\lambda+2)(\lambda+3)$，故 A 可相似对角化为

$$J=\begin{bmatrix}-1 & & \\ & -2 & \\ & & -3\end{bmatrix}$$

且由于当

$$A=\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ & & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & & 0 & 1 \\ -\alpha_n & -\alpha_{n-1} & \cdots & & -\alpha_2 & -\alpha_1 \end{bmatrix}$$

有 n 个相异特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 时，可证其对角化的相似变换矩阵

$$T=\begin{bmatrix} 1 & 1 & \cdots\cdots\cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots\cdots\cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots\cdots\cdots & \lambda_n^2 \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix}$$

即 λ_i 对应的特征向量为 $(1, \lambda_i, \lambda_i^2, \cdots, \lambda_i^{n-1})^T$，$i \in \underline{n}$，故

$$T=\begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & -3 \\ 1 & 4 & 9 \end{bmatrix}, T^{-1}=\frac{1}{2}\begin{bmatrix} 6 & 5 & 1 \\ -6 & -8 & -2 \\ 2 & 3 & 1 \end{bmatrix}$$

于是

$$e^{Jt}=\begin{bmatrix} e^{-t} & & \\ & e^{-2t} & \\ & & e^{-3t} \end{bmatrix}$$

$$e^{At}=Te^{Jt}T^{-1}$$

$$=\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & -3 \\ 1 & 4 & 9 \end{bmatrix}\begin{bmatrix} e^{-t} & & \\ & e^{-2t} & \\ & & e^{-3t} \end{bmatrix}\begin{bmatrix} 6 & 5 & 1 \\ -6 & -8 & -2 \\ 2 & 3 & 1 \end{bmatrix}$$

$$=\frac{1}{2}\begin{bmatrix} 6e^{-t}-6e^{-2t}+2e^{-3t} & 5e^{-t}-8e^{-2t}+3e^{-3t} & e^{-t}-2e^{-2t}+e^{-3t} \\ -6e^{-t}+12e^{-2t}-6e^{-3t} & -5e^{-t}+16e^{-2t}-9e^{-3t} & -e^{-t}+4e^{-2t}-3e^{-3t} \\ 6e^{-t}-24e^{-2t}+18e^{-3t} & 5e^{-t}-32e^{-2t}+27e^{-3t} & e^{-t}-8e^{-2t}+9e^{-3t} \end{bmatrix}$$

b) 已知 A 的特征值 $\lambda_1=-1$、$\lambda_2=-2$、$\lambda_3=-3$ 均为单根，故由(4)式知

$$a_1=\frac{e^{\lambda_1 t}}{\varphi_1(\lambda_1)}=\frac{e^{-t}}{(-2+1)(-3+1)}=\frac{1}{2}e^{-t}$$

同理知 $a_2=-e^{-2t}$，$a_3=\frac{1}{2}e^{-3t}$，于是

$$\begin{aligned} e^{At}&=\sum_{i=1}^{3}a_i\varphi_i(A) \\ &=\frac{1}{2}(A-\lambda_2 I)(A-\lambda_3 I)e^{-t}-(A-\lambda_1 I)(A-\lambda_3 I)e^{-2t} \\ &\quad+\frac{1}{2}(A-\lambda_1 I)(A-\lambda_2 I)e^{-3t} \end{aligned}$$

$$=\frac{1}{2}\begin{bmatrix} 6e^{-t}-6e^{-2t}+2e^{-3t} & 5e^{-t}-8e^{-2t}+3e^{-3t} & e^{-t}-2e^{-2t}+e^{-3t} \\ -6e^{-t}+12e^{-2t}-6e^{-3t} & -5e^{-t}+16e^{-2t}-9e^{-3t} & -e^{-t}+4e^{-2t}-3e^{-3t} \\ 6e^{-t}-24e^{-2t}+18e^{-3t} & 5e^{-t}-32e^{-2t}+27e^{-3t} & e^{-t}-8e^{-2t}+9e^{-3t} \end{bmatrix}$$

与 *a*)的结果一致。

§6.5 $f(A)$用有限级数表示(标准形Ⅱ)

上节介绍的 $f(A)$的拉格朗日-西勒维斯特内插多项式，$p(\lambda)$最高是 $m-1$ 次多项式($m=\deg\psi_m(\lambda)$，$\psi_m(\lambda)$为 A 的最小多项式)，将 λ 的同幂项合并得

$$p(\lambda)=a_0+a_1\lambda+a_2\lambda^2+\cdots+a_{m-1}\lambda^{m-1}$$

因此，对任意的矩阵 $A\in C^{n\times n}$，矩阵函数 $f(A)$可表示为

$$f(A)=a_0 I_n+a_1 A+a_2 A^2+\cdots+a_{m-1}A^{m-1} \quad (1)$$

其中系数 $a_0,a_1,a_2,\cdots,a_m$ 满足下面的方程组

$$\begin{cases} f(\lambda_1)=a_0+a_1\lambda_1+\cdots+a_{m-1}\lambda_1^{m-1} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ f^{(d_1-1)}(\lambda_1)=a_{(d_1-1)}(d_1-1)!+a_{d_1}(d_1!)\lambda_1+\cdots \\ \qquad +a_{m-1}\dfrac{(m-1)!}{(d_1-1)!}\lambda_1^{m-d_1} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ f(\lambda_\sigma)=a_0+a_1\lambda_\sigma+a_2\lambda_\sigma^2+\cdots+a_{m-1}\lambda_\sigma^{m-1} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ f^{(d_\sigma-1)}(\lambda_\sigma)=a_{(d_\sigma-1)}(d_\sigma-1)!+a_{d_\sigma}(d_\sigma!)\lambda_\sigma+\cdots \\ \qquad +a_{m-1}\dfrac{(m-1)!}{(d_\sigma-1)!}\lambda_\sigma^{m-d_\sigma} \end{cases}$$

特别，当 $d_i=1(i\in\underline{\sigma})$时，上面的方程组可用矩阵形式表为

$$\begin{bmatrix} f(\lambda_1) \\ f(\lambda_2) \\ \vdots \\ f(\lambda_\sigma) \end{bmatrix} = \begin{bmatrix} 1 & \lambda_1\cdots\lambda_1^{m-1} \\ 1 & \lambda_2\cdots\lambda_2^{m-1} \\ & \cdots\cdots\cdots \\ 1 & \lambda_\sigma\cdots\lambda_\sigma^{m-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{m-1} \end{bmatrix} \qquad m=\sigma$$

矩阵函数 $f(A)$的(1)式表示式，称为用有限级数表示，并简称为 $f(A)$的标准形 I 。

注意：由于 A 的特征多项式与 A 的最小多项式在 A 的谱上一致，故 $f(A)$也可用特征多项式定义为

$$f(A) = b_0I + b_1A + b_2A^2 + \cdots + b_{n-1}A^{n-1} \qquad (2)$$

其中系数 $b_0, b_1, b_2, \cdots, b_{n-1}$可由

$$q(\lambda) = b_0 + b_1\lambda + b_2\lambda^2 + \cdots + b_{n-1}\lambda^{n-1}$$

与 $f(z)$在 A 的谱上一致，即满足

$$f^{(l)}(\lambda_i) = q^{(l)}(\lambda_i) \qquad i\in\underline{\sigma},\ l=0,1,\cdots,m_i-1$$

来确定。

例1 设

$$A=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}=I_2$$

用有限级数法求 e^A。

解　易知 A 的特征多项式与最小多项式分别为

$$\psi(\lambda) = (\lambda - 1)^2, \psi_m(\lambda) = \lambda - 1$$

a）　用最小多项式。由于此时 $m=1$，有

$$p(\lambda) = a_0$$

从而

$$e^A = a_0 I_2$$

又由 $e^{\lambda_1}=p(\lambda_1)=a_0$，得$(\lambda_1=1)a_0=e$，于是

$$e^A = eI_2$$

b）　用特征多项式。由于此时 $n=2$，有

$$q(\lambda) = b_0 + b_1\lambda$$

又由

$$\begin{cases} e^{\lambda_1} = e = b_0 + b_1 \\ (e^{\lambda})' |_{\lambda=\lambda_1} = e = b_1 \end{cases}$$

故 $b_1=e, b_0=0$，于是 $e^A=eI_2$。

例 2　设

$$A = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

用有限级数法求 e^{At}。

解　由

$$\lambda I_3 - A = \begin{bmatrix} \lambda & 0 & 3 \\ 0 & 1 & 0 \\ -1 & 0 & \lambda-3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda-1 & 0 \\ 0 & 0 & (\lambda-1)(\lambda-2) \end{bmatrix}$$

有

$$\psi(\lambda) = (\lambda - 1)^2(\lambda - 2),\ \psi_m(\lambda) = (\lambda - 1)(\lambda - 2)$$

a）　用最小多项式，则

$$p(\lambda) = a_0 + a_1\lambda$$

解方程组

$$\begin{cases} e^t = a_0 + a_1 \\ e^{2t} = a_0 + 2a_1 \end{cases}$$

得 $a_0 = 2e^t - e^{2t}$，$a_1 = e^{2t} - e^t$，于是

$$e^{At} = a_0 I_3 + a_1 A$$

$$= \begin{bmatrix} 2e^t - e^{2t} & & \\ & 2e^t - e^{2t} & \\ & & 2e^t - e^{2t} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 2e^t - e^{-2t} \\ 0 & e^{2t} - e^t & 0 \\ e^{2t} - e^t & 0 & 3e^{2t} - 3e^t \end{bmatrix}$$

$$= \begin{bmatrix} 2e^t - e^{2t} & 0 & 2e^t - 2e^{2t} \\ 0 & e^t & 0 \\ e^{2t} - e^t & 0 & 2e^{2t} - e^t \end{bmatrix}$$

b) 用特征多项式，有

$$q(\lambda) = b_0 + b_1\lambda + b_2\lambda^2$$

解方程组

$$\begin{cases} e^t = b_0 + b_1 + b_2 \\ te^t = b_1 + 2b_1 \\ e^{2t} = b_0 + 2b_1 + 4b_2 \end{cases}$$

得

$$b_0 = e^{2t} - 2te^t,\ b_1 = 3te^t - e^{2t} + 2e^t,\ b_2 = e^{2t} - e^t - te^t$$

又

$$A^2 = \begin{bmatrix} -2 & 0 & -6 \\ 0 & 1 & 0 \\ 3 & 0 & 7 \end{bmatrix}$$

代入得

$$e^{At} = b_0 I_3 + b_1 A + b_2 A^2$$

$$= \begin{bmatrix} b_0 & & \\ & b_0 & \\ & & b_0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & -2b_1 \\ 0 & b_1 & 0 \\ b_1 & 0 & 3b_1 \end{bmatrix} + \begin{bmatrix} -2b_2 & 0 & -6b_2 \\ 0 & b_2 & 0 \\ 3b_2 & 0 & 7b_2 \end{bmatrix}$$

$$= \begin{bmatrix} 2e^t - e^{2t} & 0 & 2e^t - 2e^{2t} \\ 0 & e^t & 0 \\ e^{2t} - e^t & 0 & 2e^{2t} - e^t \end{bmatrix}.$$

注意:由例 2 求 a_0、a_1(b_0、b_1、b_2)的过程可知,a_0、a_1(b_0、b_1、b_2)是线性无关函数 e^t,e^{2t}(e^t、te^t、e^{2t})的线性组合.因此,a_0、a_1(b_0、b_1、b_2)也是线性无关的 t 的函数,故一般可表示为

$$f(A) = a_0(t)I_n + a_1(t)A + \cdots + a_{m-1}(t)A^{m-1}$$

或

$$f(A) = b_0(t)I_n + b_1(t)A + \cdots + b_{n-1}(t)A^{n-1}$$

其中 $a_0(t), a_1(t), \cdots, a_{m-1}(t)$($b_0(t), b_1(t), \cdots, b_{n-1}(t)$)是 t 的线性无关函数组。

习 题 六

一、已知 $A^l \neq O$, $l \in \underline{k-1}$, $A^k = O$,是否知道 A 的最小多项式及 A 的约当标准形?

二、设

$$f(A) = \sum_{i=1}^{s} \sum_{l=1}^{d_i - 1} \left(\frac{d^l f}{d\lambda^l} \right) \Bigg|_{\lambda=\lambda_i} E_i^l$$

试证 a) $E_i^l E_j^p = O \quad i \neq j \quad \forall\ l, p \in N$

b) $\sum_{i=1}^{s} E_i^0 = I_n$

三、设

$$A = \begin{bmatrix} 0 & 1 \\ -6 & -5 \end{bmatrix}$$

求矩阵函数 e^{At} 的表达式：

a) 用标准形 Ⅰ；

b) 用标准形 Ⅱ；

c) 用标准形 Ⅲ。

四、设

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

求 e^A、$\sin A$、$\cos A$、A^{100}。

五、设 A 的特征值为 λ_i，$i \in \underline{\sigma}$，则

$$\det e^A = e^{\sum_{i=1}^{\sigma} \lambda_i} = e^{\operatorname{tr} A}$$

六、设

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

试用凯莱-哈米尔顿定理求

$$A^7 - 4A^6 + 5A^5 - 2A^4 + I_3$$

第七章　广义逆矩阵

当 $A\in C_n^{n\times n}$ 时，方程组

$$Ax=b$$

的解存在、唯一，且 $x=A^{-1}b$。这一事实可否推广到一般的线性方程组

$$A_{m\times n}x=b$$

上去，使得其解 $x=B_{n\times m}b$?

1920 年摩勒(E. N. Moore)首先引进了广义逆矩阵这一概念，其后三十年未能引起人们的重视，到了 1950 年后，由于电子计算机的出现，推动了计算科学的发展，广义逆矩阵才引起了普遍的关注，从而得到迅速发展，现在它已成为代数课程的基本内容之一。这里只能介绍它的基本性质及简单的应用，从此也能了解到科技界很乐意使用这一工具的原因。

§7.1　广义逆矩阵及其性质

定义　对矩阵 $A\in C^{m\times n}$，若有矩阵 $B\in C^{n\times m}$ 使

$$ABb=b \qquad \forall\, b\in R(A) \tag{1}$$

则称 B 为 A 的广义逆矩阵，并记 $B=A^{-}$。

由(1)式知，若 $A^{-}\in C^{n\times m}$ 为 A 的广义逆矩阵，则对任意的 $b\in R(A)$，$A^{-}b$ 是方程组

$$Ax=b \tag{2}$$

的一个解。

广义逆矩阵 A^- 具有以下性质。

定理 1　$A\in C^{m\times n}$ 的广义逆矩阵 A^- 存在的充要条件为：存在满足

$$AA^-A=A \tag{3}$$

的 A^-，或说是矩阵方程

$$AXA=A \tag{4}$$

有解。

证　必要性：设有 $A^-\in C^{n\times m}$ 使对任意的 $b\in R(A)$ 有 $AA^-b=b$，则取 $b_i=\alpha_i, i\in n, A=(\alpha_1\alpha_2\cdots\cdots\alpha_n)$ 时有

$$AA^-(\alpha_1\alpha_2\cdots\cdots\alpha_n)=(\alpha_1\alpha_2\cdots\cdots\alpha_n)$$

即

$$AA^-A=A$$

充分性：设有 $A^-\in C^{n\times m}$ 使(3)式成立。由于对任意的 $b\in R(A)$ 有 $x\in C^n$ 使

$$Ax=b$$

于(3)式两端右乘 x 得

$$AA^-Ax=Ax$$

于是 $AA^-b=b$，由定义，A^- 即为 A 的广义逆矩阵。

由定理一即知

$$\operatorname{rank}A\leqslant\operatorname{rank}A^-$$

定理 2　任给 $A\in C_r^{m\times n}$，总有 $M\in C_m^{m\times m}, P\in U^{n\times n}$，使得

$$MAP=\begin{bmatrix} I_r & C \\ O & O \end{bmatrix}$$

且

$$X=P\begin{bmatrix} I_r & L_1 \\ O & L_2 \end{bmatrix}M$$

是 A 的广义逆矩阵，其中，$L_1\in C^{r\times(n-r)}$，$L_2\in C^{(n-r)\times(n-r)}$ 为任意矩阵。

证 由于 $\mathrm{rank}A=r$，故有初等变换矩阵 P，使得

$$AP=(A_1A_2)\quad A_1\in C_r^{m\times r},P\in U^{n\times n}$$

因为 $A_2\in C^{m\times(n-r)}$ 的列均能用 A_1 的列线性表示，故存在矩阵 $C\in C^{r\times(n-r)}$，使得 $A_2=A_1C$，于是

$$AP=A_1(I_rC)$$

由于 $\mathrm{rank}A_1=r$，故存在初等变换矩阵 $K_1\in U^{m\times m}$，使得

$$K_1A_1=\begin{bmatrix}A_{11}\\A_{21}\end{bmatrix}\quad A_{11}\in C_r^{r\times r}$$

由于 $A_{21}\in C^{(m-r)\times r}$ 的行均能用 A_{11} 的行线性表示，故存在矩阵 $B\in C^{(m-r)\times r}$，使得 $A_{21}=BA_{11}$，这样就有

$$K_1AP=\begin{bmatrix}I_r\\B\end{bmatrix}A_{11}(I_rC)$$

再令

$$M_1=\begin{bmatrix}A_{11}^{-1}&O\\-B&I_{n-r}\end{bmatrix},\ M=M_1K_1$$

则

$$MAP=\begin{bmatrix}A_{11}^{-1}&O\\-B&I_{n-r}\end{bmatrix}\begin{bmatrix}I_r\\B\end{bmatrix}A_{11}(I_rC)$$

$$=\begin{bmatrix}I_r&C\\O&O\end{bmatrix}$$

于是

$$A=M^{-1}\begin{bmatrix}I_r&C\\O&O\end{bmatrix}P^H$$

且

$$AXA = M^{-1}\begin{bmatrix} I_r & C \\ O & O \end{bmatrix} P^H P \begin{bmatrix} I_r & L_1 \\ O & L_2 \end{bmatrix} MM^{-1}$$

$$\begin{bmatrix} I_r & C \\ O & O \end{bmatrix} P^H = M^{-1}\begin{bmatrix} I_r & C \\ O & O \end{bmatrix} P^H = A$$

故 $X = P\begin{bmatrix} I_r & L_1 \\ O & L_2 \end{bmatrix} M$ 是 A 的一个广义逆矩阵。

如果说定理 1 给出了广义逆矩阵定义的一个等价命题，定理 2 则解决了广义逆矩阵 A^- 的存在性。

例 1 设

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

易知

$$ABA = A, \qquad ADA = A$$

即 B 与 D 均为 A 的广义逆矩阵。

定理 3 设 $A \in C^{m\times n}, \lambda \in C$，则

a) $(A^T)^- = (A^-)^T, (A^H)^- = (A^-)^H$

b) 若 $\text{rank}A = m = n$，则 $A^- = A^{-1}$ 且唯一；

c) $\lambda^+ A^-$ 是 λA 的广义逆矩阵，其中

$$\lambda^+ = \begin{cases} \dfrac{1}{\lambda} & \lambda \neq 0 \\ 0 & \lambda = 0 \end{cases}$$

d) 若 $S \in C_m^{m\times m}, T \in C_n^{n\times n}$，且 $B = SAT$，则 $T^{-1}A^-S^{-1}$ 是 B 的广义逆矩阵；

e) AA^-、A^-A 均为幂等矩阵，且

$$\mathrm{rank}A = \mathrm{rank}AA^{-} = \mathrm{rank}A^{-}A$$

f) $R(AA^{-})=R(A)$，$N(A^{-}A)=N(A)$

g) 若 $ABA=A$，$(AB)^{H}=AB$，则

$$AB = P_{R(A)}$$

证 a) 由 $AA^{-}A=A$，得

$$A^{T}(A^{-})^{T}A^{T} = A^{T}$$

由此可知$(A^{-})^{T}=(A^{T})^{-}$(这里应理解为$(A^{-})^{T}$是A^{T}的一个广义逆矩阵)；同理可证，$(A^{H})^{-}=(A^{-})^{H}$；

b) 由于此时 $A\in C_{n}^{n\times n}$可逆，且

$$A^{-1} = A^{-1}AA^{-1} = A^{-1}AA^{-}AA^{-1} = A^{-}$$

其中 A^{-} 为 A 的任意广义逆矩阵，由于 A^{-1}唯一，故 $A^{-}=A^{-1}$ 也唯一；

c)、d) 显然；

e) $(AA^{-})^{2}=(AA^{-}A)A^{-}=AA^{-}$，同理$(A^{-}A)^{2}=A^{-}A$。又

$$\mathrm{rank}A = \mathrm{rank}AA^{-}A \leqslant \mathrm{rank}AA^{-} \leqslant \mathrm{rank}A$$

故 $\mathrm{rank}AA^{-}=\mathrm{rank}A$；同理可证 $\mathrm{rank}A^{-}A=\mathrm{rank}A$；

f) 因为 $R(AA^{-})\subset R(A)$，$N(A^{-}A)\subset N(A)$，又由 e) 知

$$\dim R(A)=\dim R(AA^{-})$$
$$\dim N(A)=\dim N(A^{-}A)$$

故

$$R(AA^{-})=R(A)，N(A^{-}A)=N(A)$$

g) 由$(AB)^{H}=AB$ 及 e)，有

$$AB=(AB)^{2}=(AB)^{H}$$

故由第二章知 AB 是由 C^{n} 到 $R(AB)$的正交投影，即

$$AB=P_{R(AB)}$$

又由 f)知

$$AB=P_{R(A)}$$

推论 $A\in C^{m\times n}$，则

a) $\mathrm{rank}A=n$ 的充要条件为 $A^{-}A=I_n$；

b) $\mathrm{rank}A=m$ 的充要条件为 $AA^{-}=I_m$。

证 a) 由定理 3 之 f)知，当 $\mathrm{rank}A=n$ 时，$\mathrm{rank}A^{-}A=n$，于是 n 阶方阵 $A^{-}A$ 可逆，且

$$I_n=(A^{-}A)(A^{-}A)^{-1}=(A^{-}A)(A^{-}A)(A^{-}A)^{-1}(A^{-}A)^{-1}$$

$$=(A^{-}A)(A^{-}A)^{-1}(A^{-}A)^{-1}=(A^{-}A)^{-1}$$

故 $A^{-}A=I_n$；

反之，若 $A^{-}A=I_n$，则 $\mathrm{rank}A^{-}A=n$，所以 $\mathrm{rank}A=n$。

b) 同理可证。略。

关于 A^{-}的性质暂介绍到此。

§7.2 自反广义逆矩阵

已知当 $A\in C_n^{n\times n}$时，$(A^{-1})^{-1}=A$，这一事实对于广义逆矩阵一般不成立。例如，由上节例 1 知

$$A=\begin{bmatrix}1&0\\1&0\\1&0\end{bmatrix}$$

时，可取

$$A^{-}=\begin{bmatrix}1&0&0\\0&1&0\end{bmatrix}$$

但

$$A^{-}AA^{-}=\begin{bmatrix}1&0&0\\1&0&0\end{bmatrix}\neq A^{-}$$

即$(A^-)^- \neq A$，A与A^-能互为广义逆矩阵时，称为自反广义逆矩阵。

定义 设$A \in C^{m\times n}$，我们称使

$$AA_r^- A = A, \quad A_r^- AA_r^- = A_r^- \tag{1}$$

同时成立的$A_r^- \in C^{n\times m}$为A的**自反广义逆矩阵**（当然也可称A为A_r^-的自反广义逆矩阵）。

例1 设$A=(a_1, a_2, \cdots, a_r) \in C^{m\times r}$，且

$$a_i^H a_j = \delta_{ij} \qquad i, j \in \underline{r}$$

则A^H是A的自反广义逆矩阵。

事实上，因为

$$AA^H A = AI_r = A, \quad A^H AA^H = I_r A^H = A^H$$

故A^H是A的自反广义逆矩阵。

自反广义逆矩阵有以下性质。

定理1 设X、$Y \in C^{n\times m}$均为$A \in C^{m\times n}$的广义逆矩阵，则

$$Z = XAY \tag{2}$$

是A的自反广义逆矩阵。

证 因为

$$AZA=(AXA)YA=AYA=A$$

$$ZAZ=X(AYA)XAY=X(AXA)Y=XAY=Z$$

所以$Z=XAY$是A的自反广义逆矩阵。

这个定理给出了自反广义逆矩阵的一种具体构造方法。

定理2 设$A^- \in C^{n\times m}$是$A \in C^{m\times n}$的广义逆矩阵，则A^-是A的自反广义逆矩阵的充要条件为

$$\operatorname{rank} A^- = \operatorname{rank} A \tag{3}$$

证 必要性：设A^-是A的自反广义逆矩阵，则

$$AA^- A = A, \quad A^- AA^- = A^-$$

于是由

$$\mathrm{rank}A \leqslant \mathrm{rank}A^-, \ \mathrm{rank}A^- \leqslant \mathrm{rank}A$$

知(3)式成立；

充分性：设 $AA^-A=A$，且 $\mathrm{rank}A=\mathrm{rank}A^-$，于是由 $R(A^-A)\subset R(A^-)$ 及 $\mathrm{rank}A^-A=\mathrm{rank}A^-$，有

$$R(A^-A)=R(A^-)$$

故存在 $X\in C^{n\times m}$，使得

$$A^-=A^-AX$$

且

$$A=AA^-A=(AA^-A)XA=AXA$$

故 X 也是 A 的广义逆矩阵。于是由定理1知 A^- 是 A 的自反广义逆矩阵。

这个定理给出了在广义逆矩阵中，区分自反广义逆矩阵的一种有效方法，当然，也可认为这是求自反广义逆矩阵的具体方法。

定理 3 设 $A\in C^{m\times n}$，$X\in C^{n\times m}$，则

a) $\mathrm{rank}A=\mathrm{rank}X$

b) $AXA=A$

c) $XAX=X$

中任两条成立，可推得第三条成立。

证明留给读者自己练习。

定理 4 设 $A\in C^{m\times n}$，则

$$Y=(A^HA)^-A^H, \ Z=A^H(AA^H)^- \tag{4}$$

均是 A 的自反广义逆矩阵。

证 由于对任意的 $A\in C^{m\times n}$，总有

$$R(A^H)=R(A^HA), \ N(A)=N(A^HA)$$

故存在矩阵 $D\in C^{n\times m}$，使得

$$A^H=A^HAD$$

则

$$AYA = D^H A^H AYA = D^H A^H A(A^H A)^- A^H A$$
$$= D^H A^H A = A$$

且由 $\mathrm{rank}Y \leqslant \mathrm{rank}A^H$ 及

$$A^H A(A^H A)^- A^H A = A^H AYA = A^H A$$

有 $\mathrm{rank}Y \geqslant \mathrm{rank}A^H A = \mathrm{rank}A^H$，于是

$$\mathrm{rank}Y = \mathrm{rank}A^H = \mathrm{rank}A$$

故由定理 3 知，Y 是 A 的自反广义逆矩阵；

同理可证 Z 是 A 的自反广义逆矩阵。

定理 5 设 $A=BC$ 为矩阵 A 的最大秩分解，则 A 的自反广义逆矩阵的一般形式为

$$A_r^- = C_R^{-1} B_L^{-1} \tag{5}$$

其中 $B_L^{-1}B = CC_R^{-1} = I_r$（称 B_L^{-1} 为 B 的**左逆矩阵**，称 C_R^{-1} 为 C 的**右逆矩阵**）。

证 由于

$$AA_r^- A = BCC_R^{-1}B_L^{-1}BC = BC = A$$
$$A_r^- AA_r^- = C_R^{-1}B_L^{-1}BCC_R^{-1}B_L^{-1} = C_R^{-1}B_L^{-1} = A_r^-$$

故 $A_r^- = C_R^{-1}B_L^{-1}$ 是 A 的自反广义逆矩阵；

其次，设 A_r^- 是 A 的任一自反广义逆矩阵，则

$$AA_r^- A = A$$

即

$$BCA_r^- BC = BC$$

上式两端分别左乘以 B_L^{-1}，右乘以 C_R^{-1}，得

$$CA_r^- B = I_r \qquad (r = \mathrm{rank}A)$$

由此可见，CA_r^{-1} 为 B 的左逆矩阵，记为 $\tilde{B}_L^{-1}$，$A_r^- B$ 为 C 的右逆矩阵，记为 $\tilde{C}_R^{-1}$，于是

$$A_r^- = A_r^- AA_r^- = A_r^- BCA_r^- = \tilde{C}_R^{-1}\tilde{B}_L^{-1}$$

故(5)式是 A 的自反广义逆矩阵的一般形式。

注意：若 $A=BC$ 是 A 的最大秩分解，显然 $(B^HB)^{-1}B^H$ 是 B 的左逆矩阵，$C^H(CC^H)^{-1}$ 是 C 的右逆矩阵。这样一来，(5)式也给出了求 A 的自反广义逆矩阵的一种方法。

§7.3 伪逆矩阵

在自反广义逆矩阵中，还有一类更特殊也更为重要的广义逆矩阵，这就是将要介绍的伪逆矩阵。它不仅在应用上特别重要，而且有很多有趣的性质。

定义 设 $A\in C^{m\times n}$，$X\in C^{n\times m}$ 若同时有

$$\begin{cases} AXA=X \\ XAX=X \\ (AX)^H=AX \\ (XA)^H=XA \end{cases} \tag{1}$$

则称 X 为 A 的**伪逆矩阵**，记为

$$X=A^+$$

方程组(1)称为**彭罗司-摩勒(Penrose-Moore)方程**。从定义中看到，在方程组(1)中，X 与 A 完全处于对称地位，因此 A 也是 X 的伪逆矩阵。

定理1 若 $A\in C^{m\times n}$，且 $A=BC$ 是最大秩分解，则

$$X=C^H(CC^H)^{-1}(B^HB)^{-1}B^H \tag{2}$$

是 A 的伪逆矩阵。

证
$$\begin{aligned} AXA &=BCC^H(CC^H)^{-1}(B^HB)^{-1}B^HBC \\ &=BC=A \end{aligned}$$

$$\begin{aligned} XAX &=C^H(CC^H)^{-1}(B^HB)^{-1}B^HBCC^H(CC^H)^{-1}(B^HB)^{-1}B^H \\ &=C^H(CC^H)^{-1}(B^HB)^{-1}B^H=X \end{aligned}$$

$$(AX)^H=[BCC^H(CC^H)^{-1}(B^HB)^{-1}B^H]^H$$

$$=[B(B^HB)^{-1}B^H]^H=B(B^HB)^{-1}B^H=AX$$

$$(XA)^H=[C^H(CC^H)^{-1}(B^HB)^{-1}(B^HBC)]^H$$

$$=[C^H(CC^H)^{-1}C]^H=C^H(CC^H)^{-1}C=XA$$

这个定理不仅证明了伪逆矩阵的存在，而且也给出了求伪逆矩阵的具体办法。

推论 设 $A\in C_r^{m\times n}$，则

a) 当 $r=n$ 时

$$A^+=(A^HA)^{-1}A^H \tag{3}$$

b) 当 $r=m$ 时

$$A^+=A^H(AA^H)^{-1} \tag{4}$$

这只要注意到 $A=AI_n=I_mA$，即知。

定理 2 伪逆矩阵唯一。

证 设 X,Y 均为 A 的伪逆矩阵，则

$$XAX=XAYAX=XY^HA^HX^HA^H=XY^HA^H$$

$$=XAY=A^HX^HY=A^HY^HA^HX^HY=YAXAY$$

$$=YAY=Y$$

注意：在唯一性的证明里，对 X 与 Y 均用到了方程组(1)中的四个方程。

伪逆矩阵具有以下性质。

定理 3 设 $A\in C^{m\times n}$，则

a) $(A^T)^+=(A^+)^T$，$(A^H)^+=(A^+)^H$

b) $(A^+)^+=A$

c) $A^+=A^H(AA^H)^+=(A^HA)^+A^H$

d) $R(A^+)=R(A^H)$

e) $AA^+=P_{R(A)}$，$A^+A=R_{R(A^H)}$

f) $R(A)=R(A^H)$的充要条件为

$$AA^+=A^+A$$

证 由定义，a)、b)显然成立；

c) 由上节定理4知，$A^H(AA^H)^+$、$(A^HA)^+A^H$ 均为 A 的自反广义逆矩阵，又

$$A^H(AA^H)^+A=[A^H(AA^H)^+A]^H$$
$$AA^H(AA^H)^+=[AA^H(AA^H)^+]^H$$
$$(A^HA)^+A^HA=[(A^HA)^+A^HA]^H$$
$$A(A^HA)^+A^H=[A(A^HA)^+A^H]^H$$

故 $A^H(AA^H)^+$，$(A^HA)^+A^H$ 是 A 的伪逆矩阵，即

$$A^+=A^H(AA^H)^+=(A^HA)^+A^H$$

d) 由 $\mathrm{rank}A=\mathrm{rank}A^H=\mathrm{rank}A^+$ 及 c)知

$$R(A^+)=R(A^H)$$

e) 由$(AA^+)^2=AA^+$、$(AA^+)^H=AA^+$知，AA^+是正交投影，再由§7.1定理3之g)得

$$AA^+=P_{R(AA^+)}=P_{R(A)}$$

同理有

$$A^+A=P_{R(A^+A)}=P_{R(A^+)}=P_{R(A^H)}$$

f) 若 $R(A)=R(A^H)$，则由 e)有

$$AA^+=P_{R(A)}=P_{R(A^H)}=A^+A$$

反之，若 $A^+A=AA^+$，则

$$R(A)=R(AA^+)=R(A^+A)=R(A^+)=R(A^H)$$

定理4 设矩阵 A、B 可乘，则

$$(AB)^+=B^+A^+ \tag{5}$$

的充要条件为

$$R(A^HAB)\subset R(B)、R(BB^HA^H)\subset R(A^H) \tag{6}$$

证 必要性：设(5)式成立，则由

$$\begin{aligned}B^HA^H&=(AB)^H=(AB)^H[(AB)^H]^+(AB)^H\\&=(AB)^+(AB)(AB)^H=B^+A^+ABB^HA^H\end{aligned}$$

在上式两边左乘以 ABB^HB，得

$$ABB^HBB^HA^H=ABB^HBB^+A^+ABB^HA^H$$
$$=ABB^HB(B^HB)^+B^HA^+ABB^HA^H$$
$$=ABB^H(B^H)^+B^HA^+ABB^HA^H$$
$$=ABB^HA^+ABB^HA^H$$

即

$$ABB^H(I-A^+A)BB^HA^H=O$$

由于

$$I-A^+A=(I-A^+A)^2=(I-A^+A)^H$$

则得

$$[(I-A^+A)BB^HA^H]^H[(I-A^+A)BB^HA^H]=O$$

从而有

$$(I-A^+A)BB^HA^H=O$$

即

$$BB^HA^H=A^+ABB^HA^H$$

于是

$$R(BB^HA^H)=R(A^+ABB^HA^H)$$
$$\subset R(A^+A)=R(A^H)$$

同理可证

$$R(A^HAB)\subset R(B)$$

即(6)式成立；

充分性：由于

$$BB^+=P_{R(B)}, \qquad A^+A=P_{R(A^H)}$$

及(6)式成立，有

$$BB^+A^HAB=A^HAB, \quad A^+ABB^HA^H=BB^HA^H \tag{7}$$

由第一式得

$$B^HA^HA=(AB)^HABB^+$$

上式两边左乘以$[(AB)^H]^+$，右乘以 A^+得

$$[(AB)^H]^+(AB)^H AA^+ = [(AB)^H]^+(AB)^H(AB)B^+A^+$$
$$= \{[(AB)^H]^+(AB)^H\}^H(AB)B^+A^+ = ABB^+A^+$$

即

$$AB(AB)^+AA^+ = ABB^+A^+$$

亦即

$$P_{R(AB)}P_{R(A)} = ABB^+A^+$$

又

$$P_{R(AB)}P_{R(A)} = P_{R(AB)}$$

于是

$$(AB)(AB)^+ = ABB^+A^+ \tag{8}$$

上式两边右乘 AB 得

$$AB = ABB^+A^+AB \tag{9}$$

于是

$$\mathrm{rank}AB \leqslant \mathrm{rank}B^+A^+$$

又由于

$$A^+ = A^H(AA^H)^+, \quad B^+ = (B^HB)^+B^H$$

得

$$B^+A^+ = (B^HB)^+B^HA^H(AA^H)^+$$

于是

$$\mathrm{rank}B^+A^+ \leqslant \mathrm{rank}(AB)^H = \mathrm{rank}AB$$

故

$$\mathrm{rank}B^+A^+ = \mathrm{rank}AB$$

所以

$$B^+A^+(AB)B^+A^+ = B^+A^+ \tag{10}$$

再由(7)式第二式左乘 B^+右乘$[(AB)^H]^+$得

$$B^+A^+AB(AB)^H[(AB)^H]^+ = B^+B(AB)^H[(AB)^H]^+$$

即

$$B^+A^+AB=B^+B(AB)^+AB=(AB)^+AB \tag{11}$$

于是由(8)～(11)知 B^+A^+ 是 AB 的伪逆矩阵，即(5)式成立。

定理 5 设 $A\in C^{m\times n}$，$B\in C^{n\times p}$，若

$$B_1\triangleq A^+AB, \qquad A_1\triangleq AB_1B_1^+$$

则

$$(AB)^+=B_1^+A_1^+$$

证 因为

$$A_1B_1=(AB_1B_1^+)B_1=AB_1=A(A^+AB)=AB$$

故只须证 $B_1^+A_1^+$ 是 A_1B_1 的伪逆矩阵即可。

记 $X=B_1^+A_1^+$，$Y=A_1B_1$，则

$$\begin{aligned}
XYX &=B_1^+A_1^+A_1B_1B_1^+A_1^+=B_1^+A_1^+AB_1B_1^+A_1^+\\
&=B_1^+A_1^+A_1A_1^+=B_1^+A_1^+=X\\
YXY &=A_1B_1B_1^+A_1^+A_1B_1=AB_1B_1^+A_1^+A_1B_1\\
&=A_1A_1^+A_1B_1=A_1B_1=Y\\
XY &=B_1^+A_1^+A_1B_1=B_1^+B_1B_1^+A_1^+A_1B_1\\
&=B_1^+A^+ABB_1^+A_1^+A_1B_1=B_1^+A^+AB_1B_1^+A_1^+A_1B_1\\
&=B_1^+A^+A_1A_1^+A_1B_1=B_1^+A^+A_1B_1=B_1^+A^+AB\\
&=B_1^+B_1\\
YX &=A_1B_1B_1^+A_1^+=AB_1B_1^+A_1^+=A_1A_1^+
\end{aligned}$$

由此可知，X 是 Y 的伪逆矩阵，亦即

$$(AB)^+=B_1^+A_1^+$$

注意：一般情况下，$(AB)^+\neq B^+A^+$。例如，取

$$A=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},\ B=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$$

则

$$AB=\begin{bmatrix}1 & 1\\0 & 0\end{bmatrix}$$

$$A^+=\left[\begin{pmatrix}1\\0\end{pmatrix}(1 \quad 0)\right]^+=\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$$

$$B^+=B^{-1}=\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}$$

$$(AB)^+=\left[\begin{pmatrix}1\\0\end{pmatrix}(1 \quad 1)\right]^+=\frac{1}{2}\begin{bmatrix}1 & 0\\1 & 0\end{bmatrix}$$

但

$$B^+A^+=\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}=\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\neq(AB)^+$$

§7.4 伪逆矩阵的其它表示式

由于伪逆矩阵具有很多重要性质，因此，应用十分广泛。为了适应各种场合的不同需要，寻求 A^+ 的各种表示式，就显得特别必要。下面将给出 A^+ 的几种其它常用的表示式。

定理 1 设 $A\in C^{m\times n}$，

当 $A=BC$ 为最大秩分解时，则

$$A^+=C^+B^+ \tag{1}$$

这个定理成立是显然的。

定理 2 设 $A\in C^{m\times n}$ 的奇值分解为

$$A=UDV^H$$

其中 $U\in U^{m\times m}$，$V\in U^{n\times n}$

$$D=\begin{bmatrix}\Sigma & O\\O & O\end{bmatrix}$$

$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$，$r = \mathrm{rank}\, A$，$\sigma_i\ (i \in \underline{r})$ 为 A 的奇值，则

a) $A^+ = VD^+U^H$ (2)

b) $\| A^+ \|_F^2 = \sum_{i=1}^{r} \left(\frac{1}{\sigma_i}\right)^2$

c) $\| A^+ \|_2 = \frac{1}{\sigma_0}$，$\sigma_0 = \min_i \sigma_i$

证 a) 若令

$$X = \begin{bmatrix} \Sigma^{-1} & O \\ O & O \end{bmatrix}$$

易知它满足

$$DXD = D,\ XDX = X$$

$$(DX)^H = DX,\ (XD)^H = XD$$

故 $X = D^+$，即

$$D^+ = \begin{bmatrix} \Sigma^{-1} & O \\ O & O \end{bmatrix}$$

由于

$$U^H U D D^H = D^2,\ D D^H U^H U = D^2$$

故

$$R(U^H U D D^H) = R(D D^H U^H U) = R(D^2)$$

于是

$$(UD)^+ = D^+ U^+ = D^+ U^H$$

又

$$\begin{aligned} A^+ &= A^H (AA^H)^+ = VDU^H (UDV^H VDU^H)^+ \\ &= VDU^H (UDDU^H)^+ = V(UD)^H [(UD)(UD)^H]^+ \\ &= V(UD)^+ = VD^+U^H \end{aligned}$$

即(2)式成立；

b)　由 a)有

$$A^+(A^+)^H = VD^+U^H(VD^+U^H)^H$$
$$= VD^+U^HUD^+V^H = V(D^+)^2V^H$$

且

$$(D^+)^2 = \begin{bmatrix} (\Sigma^{-1})^2 & O \\ O & O \end{bmatrix} = \begin{bmatrix} \Delta^{-1} & O \\ O & O \end{bmatrix}$$

其中 $\Delta^{-1} = \mathrm{diag}\left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \cdots, \frac{1}{\sigma_r^2}\right)$由此可知，$A^+$的正奇值为$\frac{1}{\sigma_i}$，$i \in \underline{r}$，于是

$$\| A^+ \|_F^2 = \mathrm{tr} A^+ (A^+)^H = \sum_{i=1}^{r} \frac{1}{\sigma_i^2}$$

c)　$$\| A^+ \|_2 = \max_i \sqrt{\lambda_i[A^+(A^+)^H]}$$

$$= \max_i \frac{1}{\sigma_i} = \frac{1}{\sigma_0} \qquad \sigma_0 = \min_i \sigma_i \quad i \in \underline{r}$$

定理 3　设 $A \in C^{m \times n}$，则下列条件等价：

a)　A 是次酉矩阵；

b)　$A^HA = P_{R(A^H)}$

c)　$AA^H = P_{R(A)}$

d)　$AA^HA = A$

e)　$A^HAA^H = A^H$

f)　$A^+ = A^H$，$(A^H)^+ = A$

g)　A^H 是次酉矩阵(或 A^+是次酉矩阵)。

证　a) ⇒ b) 设 $H = P_{R(A^H)} - A^HA$，则 $H^H = H$，又设 $x \in C^n$，则 $x = x_1 + x_2$，其中 $x_1 \in N(A)$，$x_2 \in R(A^H)$，于是

$x_1^H x_2=0$，且

$$x^H Hx=(x_1^H+x_2^H)(P_{R(A^H)}-A^H A)(x_1+x_2)$$
$$=(x_1^H+x_2^H)(x_2-A^H Ax_2)=x_2^H x_2-x_2^H A^H Ax_2$$
$$=\|x_2\|_2-\|Ax_2\|_2$$

由 a)知 $\|x_2\|_2=\|Ax_2\|_2$，故

$$x^H Hx=0 \qquad \forall\ x\in C^n$$

由 x 的任意性知 $H=O$，即 b)成立；

同理 $a)\Rightarrow c)$；

$b)\Rightarrow d)$　由 b)及§7.3 定理 3 之 c)有

$$A^H A=P_{R(A^H)}=A^+A$$

故

$$AA^H A=A$$

$d)\Longleftrightarrow e)$显然；

$e)\Rightarrow f)$　由 d)、e)及 $A^H A$、AA^H 均为埃尔米特矩阵，故 A^H 是 A 的伪逆矩阵，即 $A^H=A^+$，由此可知$(A^H)^+=A$；

$f)\Rightarrow g)$　对任意的 $y\in R(A)$，有

$$y=AA^+y$$

于是由 f)

$$\|y\|_2^2=y^H y=y^H AA^+y=y^H AA^H y=\|A^H y\|_2^2$$

即

$$\|A^H y\|_2=\|y\|_2 \qquad \forall\ y\in R(A)$$

故 A^H 是次酉矩阵(A^+是次酉矩阵)。

类似可证定理的其余部分，以下略。

定理 4　设 $A\in C_r^{m\times n}$，$A^H A=\sum_{i=1}^{r}\lambda_i(A^H A)E_i$ 为其谱分解，则

$$A^{+}=\sum_{i=1}^{r}\lambda_i^{-1}(A^HA)\frac{\prod\limits_{\substack{j=1\\j\neq i}}^{r}(A^HA-\lambda_jI)}{\prod\limits_{\substack{j=1\\j\neq i}}^{r}(\lambda_i-\lambda_j)}A^H$$

证 由 A^HA 的谱分解式易知

$$(A^HA)^{+}=\sum_{i=1}^{r}\lambda_i^{-1}(A^HA)E_i$$

记

$$p_i(\lambda)=\prod_{\substack{j=1\\j\neq i}}^{r}(\lambda-\lambda_j)$$

则

$$p_i(A^HA)=\sum_{j=1}^{r}p_i(\lambda_j)E_j=p_i(\lambda_i)E_i$$

由于 $p_i(\lambda_i)\neq 0$，故有

$$E_i=\frac{p_i(A^HA)}{p_i(\lambda_i)}$$

于是

$$A^{+}=(A^HA)^{+}A^H=\sum_{i=1}^{r}\lambda_i^{-1}(A^HA)E_iA^H$$

$$=\sum_{i=1}^{r}\lambda_i^{-1}(A^HA)\frac{\prod\limits_{\substack{j=1\\j\neq i}}^{r}(A^HA-\lambda_jI)}{\prod\limits_{\substack{j=1\\j\neq i}}^{r}(\lambda_i-\lambda_j)}A^H$$

下面为了给出A^+的极限形式的表示式，先证明一个引理。

引理 设 $A=A^H\in C^{n\times n}$，则

$$AA^{+}=\lim_{\delta\to 0}(A+\delta I)^{-1}A$$

$$= \lim_{\delta \to 0} A(A + \delta I)^{-1} (= A^{+} A)$$

证 对于足够小的 $|\delta| > 0$，因 $A + \delta I$ 是可逆的，故 $(A + \delta I)^{-1}$ 存在。

设 $A = U \Lambda U^{H}$，$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$，则

由

$$A = AA^{+} A = A^{2} A^{+}$$

有

$$\begin{aligned}(A + \delta I)^{-1} A &= (A + \delta I)^{-1} A^{2} A^{+} \\ &= U(\Lambda + \delta I)^{-1} U^{H} U \Lambda^{2} U^{H} A^{+} \\ &= U(\Lambda + \delta I)^{-1} \Lambda^{2} U^{H} A^{+}\end{aligned}$$

于是

$$\begin{aligned}&\lim_{\delta \to 0} (A + \delta I)^{-1} A = \lim_{\delta \to 0} [U(\Lambda + \delta I)^{-1} \Lambda^{2} U^{H} A^{+}] \\ &= U \lim_{\delta \to 0} \mathrm{diag}[(\lambda_1 + \delta)^{-1} \lambda_1^{2}, (\lambda_2 + \delta)^{-1} \lambda_2^{2}, \cdots, \\ &\qquad (\lambda_n + \delta)^{-1} \lambda_n^{2}] U^{H} A^{+} \\ &U \Lambda U^{H} A^{+} = AA^{+}\end{aligned}$$

同理可证

$$\lim_{\delta \to 0} A(A + \delta I)^{-1} = AA^{+}$$

定理 5 设 $A \in C^{m \times n}$，则

$$\begin{aligned}A^{+} &= \lim_{\delta \to 0} (A^{H} A + \delta I_n)^{-1} A^{H} \\ &= \lim_{\delta \to 0} A^{H} (AA^{H} + \delta I_m)^{-1}\end{aligned}$$

证 由引理有

$$\begin{aligned}&\lim_{\delta \to 0} (A^{H} A + \delta I_n)^{-1} A^{H} \\ &= \lim_{\delta \to 0} (A^{H} A + \delta I_n)^{-1} A^{H} AA^{+}\end{aligned}$$

$$= A^H A(A^H A)^+ A^+ = [(A^H A)^+]^H A^H A A^+$$

$$= (A^H A)^+ A^H = A^+$$

同理可证

$$A^+ = \lim_{\delta \to 0} A^H (AA^H + \delta I_m)^{-1}$$

例 设

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

求其伪逆矩阵 A^+。

解 首先对 A 进行最大秩分解

$$A \to \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

所以 A 的最大秩分解为

$$A = BC = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

由定理 1 知 $A^+ = C^+ B^+$，这里 B 为三阶可逆方阵，故

$$B^{+}=B^{-1}=\begin{bmatrix}\frac{1}{2} & -\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

又 $C^{+}=C^{H}(CC^{H})^{+}=C^{H}(CC^{H})^{-1}$，于是

$$C^{+}=C^{H}(CC^{H})^{-1}$$

$$=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}\frac{2}{3} & 0 & 0\\ 0 & \frac{3}{4} & -\frac{1}{4}\\ 0 & -\frac{1}{4} & \frac{3}{4}\end{bmatrix}$$

$$=\begin{bmatrix}\frac{2}{3} & 0 & 0\\ 0 & \frac{3}{4} & -\frac{1}{4}\\ 0 & -\frac{1}{4} & \frac{3}{4}\\ \frac{1}{3} & \frac{1}{4} & \frac{1}{4}\\ -\frac{1}{3} & \frac{1}{4} & \frac{1}{4}\end{bmatrix}$$

则

$$A^{+}=C^{+}B^{+}=\begin{bmatrix}\frac{2}{3} & 0 & 0\\ 0 & \frac{3}{4} & -\frac{1}{4}\\ 0 & -\frac{1}{4} & \frac{3}{4}\\ \frac{1}{3} & \frac{1}{4} & \frac{1}{4}\\ -\frac{1}{3} & \frac{1}{4} & \frac{1}{4}\end{bmatrix}\begin{bmatrix}\frac{1}{2} & -\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

$$=\begin{bmatrix}\frac{2}{3} & -\frac{1}{3} & \frac{1}{3}\\ \frac{1}{2} & \frac{1}{4} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{4} & \frac{1}{2}\\ \frac{1}{6} & \frac{1}{12} & \frac{1}{6}\\ -\frac{1}{6} & \frac{5}{12} & -\frac{1}{6}\end{bmatrix}$$

对于分块矩阵的 A_r^- 与 A^+ 可按下面定理求得。

定理 6 设 $A\in C_r^{m\times n}$，则存在换行，换列矩阵之积 $P\in U^{m\times m}$，$Q\in U^{n\times n}$ 与 $S\in C^{(n-r)\times r}$，$T\in C^{r\times(n-r)}$，使得

$$A=P^H\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}Q^H$$

$$=P^H\begin{bmatrix} I_r \\ S \end{bmatrix} A_{11}(I_r\ T)Q^H \quad A_{11}\in C_r^{r\times r}$$

则

$$A_r^-=Q\begin{bmatrix} A_{11}^{-1} & O \\ O & O \end{bmatrix}P$$

$$A^+=Q\begin{bmatrix} I_r \\ T^H \end{bmatrix}(I_r+TT^H)^{-1}A_{11}^{-1}(I_r+S^HS)^{-1}(I_r S)P$$

证略。

§7.5 广义逆矩阵的应用

本节主要介绍广义逆矩阵在矩阵方程及线性方程组中的简单应用。

定理 1 设 $A\in C^{m\times n}$、$B\in C^{p\times q}$、$D\in C^{m\times q}$，则矩阵方程

$$AXB=D \tag{1}$$

有解的充要条件为，存在 A^-、B^-，使得

$$AA^-DB^-B=D \tag{2}$$

且其通解可写成

$$X=A^-DB^-+Y-A^-AYBB^- \quad \forall\, Y\in C^{n\times p} \tag{3}$$

证 必要性：设(1)有解 X，则在(1)式两边左乘以 AA^-，右乘以 B^-B，得

$$AA^-AXBB^-B=AXB=D=AA^-DB^-B$$

充分性：设有 A^-、B^-使(2)式成立，则 A^-DB^-是(1)的解。

现设 X_0 是(1)的解，令 $Y=X_0$ 就有

$$A^-DB^-+X_0-A^-AX_0BB^-$$

$$=A^{-}DB^{-}+X_0-A^{-}DB^{-}=X_0$$

故(3)是(1)的通解。

推论 a) 设 A^{-}是 A 的一个广义逆矩阵，则 A 的广义逆矩阵 X 的一般形式为

$$X=A^{-}+Z-A^{-}AZAA^{-} \quad \forall Z\in C^{n\times m}$$

b) $AX=D$ 有解的充要条件为存在 A^{-}，使得

$$AA^{-}D=D$$

且其通解为

$$X=A^{-}D+Y-A^{-}AY \qquad \forall Y\in C^{n\times q}$$

其中 $A\in C^{m\times n}$，$D\in C^{m\times q}$。

证 a) 在(1)中，令 $A=B=D$，(3)中，令 $Y=A^{-}+Z$ 即得；

b) 在(1)中，令 $B=I_q$ 即得。

定理 2 设 $A\in C^{m\times n}$，$\boldsymbol{b}\in C^{m}$，则方程组

$$A\boldsymbol{x}=\boldsymbol{b} \tag{4}$$

有解的充要条件为，存在 A^{-}，使得

$$AA^{-}\boldsymbol{b}=\boldsymbol{b} \tag{5}$$

且其通解为

$$\boldsymbol{x}=A^{-}\boldsymbol{b}+(I_n-A^{-}A)\boldsymbol{y} \quad \forall \boldsymbol{y}\in C^{n} \tag{6}$$

证 在定理 1 中，令 $B=I_n$，$D=\boldsymbol{b}$ 即得。

推论 $A\boldsymbol{x}=\boldsymbol{\theta}$ 的通解为

$$\boldsymbol{x}=(I_n-A^{-}A)\boldsymbol{y} \qquad \forall \boldsymbol{y}\in C^{n}$$

此推论说明，$I_n-A^{-}A$ 是 $N(A)$上的正交投影。

定理 3 设 $A\in C^{m\times n}$、$B\in C^{m\times l}$、$D\in C^{l\times k}$、$E\in C^{n\times k}$，则

$$AX=B,\ XD=E \tag{7}$$

有公共解的充要条件为，两个方程分别有解，且

$$AE=BD \tag{8}$$

若 X_0 为(7)的解，则其通解为

$$X=X_0+(I_n-A^-A)Y(I_l-DD^-) \quad \forall Y\in C^{n\times l} \tag{9}$$

证 必要性：设(7)有公共解 X，于是

$$AXD=BD, \ AXD=AE$$

即(8)式成立；

充分性：设(7)式的两个方程分别有解，且(8)式成立，则令

$$X=A^-B+ED^--A^-AED^-$$

易验证 X 同时满足(7)的两个方程，即(7)的两个方程有公共解。

若 X_0 是(7)的一个公共解，则对任何 $Y\in C^{n\times l}$，由通解式(9)确定的 X，满足(7)的两个方程。

设 X 是(7)的公共解，则

$$A(X-X_0)=O, \ (X-X_0)D=O$$

由此有

$$R(X-X_0)\subset N(A) \quad R(D)\subset N(X-X_0)$$

于是存在矩阵 Y、Z，使得

$$X-X_0=(I_n-A^-A)Y=Z(I_l-DD^-)$$

故

$$X-X_0=(I_n-A^-A)^2Y=(I_n-A^-A)Z(I_l-DD^-)$$

即(9)式成立。

定理 4 设 A，$B\in C^{m\times n}$，则

$$Ax=a, \ Bx=b \tag{10}$$

有公共解的充要条件为

$$B^+b-A^+a\in N(A)+N(B) \tag{11}$$

证 必要性：设(10)有公共解，由定理 2 有

$$x=A^{+}a+(I_n-A^{-}A)y \qquad \forall\ y\in C^n$$
$$x=B^{+}b+(I_n-B^{-}B)z \qquad \forall\ z\in C^n$$

于是

$$B^{+}b-A^{+}a=(I_n-A^{-}A)y-(I_n-B^{-}B)z\in N(A)+N(B)$$

充分性：设(11)式成立，即

$$B^{+}b-A^{+}a=\tilde{a}+\tilde{b} \qquad \tilde{a}\in N(A),\tilde{b}\in N(B)$$

令

$$x=B^{+}b-\tilde{b}=A^{+}a+\tilde{a}$$

则

$$Ax=AA^{+}a+A\tilde{a}=a$$
$$Bx=BB^{+}b-B\tilde{b}=b$$

即 x 是(10)的公共解。

从以上讨论可知，方程有解时，往往有多个解。而在实践中往往要求在这多个解中，寻求具有某种条件的解。下面介绍其中范数最小的解。

定义 设 $A\in C^{m\times n}$、$b\in C^m$，方程

$$Ax=b \tag{4}$$

有解时，解中范数最小的称为**最小范数解**，简记为 LN 解，而当 $b\overline{\in}R(A)$时，求使

$$\|Ax-b\|_2=\min \tag{12}$$

的解中范数最小的解，称为**最小二乘解问题的最小范数的最小二乘解**，简记为 LNLS 解。

定理5 方程(4)有解时，其 LN 解存在且唯一，若 $x_0\in C^n$是其 LN 解，则

$$Ax_0=b \qquad x_0\in R(A^H)$$

证 由于 $A: R(A^H)\to R(A)$ 是线性映射，且 $R(A^H)\cap N(A)=\{\theta\}$，故 A 是由 $R(A^H)$ 到 $R(A)$ 的同构映射。

若(4)有解，则 $b\in R(A)$，于是存在唯一的 $x_0\in R(A^H)$，使得

$$Ax_0=b \qquad x_0\in R(A^H)$$

对于方程(4)的任意解 $x_1\in C^n$，均有

$$x_1-x_0\in N(A)$$

由于 $R(A^H)\perp N(A)$，有

$$\begin{aligned}\|x_1\|_2^2&=\|(x_1-x_0)+x_0\|_2^2\\&=\|x_1-x_0\|_2^2+\|x_0\|_2^2\geqslant\|x_0\|_2^2\end{aligned}$$

故 x_0 是 LN 解。

定理6 若方程(4)有解，则其唯一的 LN 解为

$$x=Zb \tag{13}$$

其中 Z 满足

$$AZA=A \qquad (ZA)^H=ZA \tag{14}$$

反之，对一切 $b\in R(A)$，$x=Zb$ 都是(4)对应的 LN 解，则(14)式成立。

证 若(14)式成立，且方程(4)有解，则 $b\in R(A)$，且存在 $c\in C^n$，使得

$$Ac=b$$

于是

$$AZb=AZAc=Ac=b$$

另一方面

$$Zb=ZAc=(ZA)^Hc=A^HZ^Hc\in R(A^H)$$

于是由定理5知，$Zb=x_0$ 是其唯一的 LN 解；

反之，令 $A=(a_1, a_2, \cdots, a_n)$，$a_i\in R(A)$，$i\in \underline{n}$，则

$$AZa_i=a_i \qquad Za_i\in R(A^H) \qquad i\in \underline{n}$$

由此得

$$AZA=A$$

$$ZA=A^+AZA=A^+A$$

于是

$$(ZA)^H=ZA$$

证毕。

定理7 设 $A\in C^{m\times n}$，$b=C^m$，则当

$$x=A^-b \qquad (AA^-A=A, (AA^-)^H=AA^-) \qquad (15)$$

时，(15)为方程

$$\|Ax-b\|_2=\min$$

的解。

反之，若 $X\in C^{n\times m}$能对一切 $b\in C^m$有

$$\|AXb-b\|_2=\min$$

则

$$AXA=A \qquad (AX)^H=AX$$

证 对任意的 $b\in C^m$，令

$$b=b_1+b_2$$

其中 $b_1=P_{R(A)}b\in R(A)$，$b_2=(I-P_{R(A)})b\in R(A)_\perp$，则

$$b_1\perp b_2$$

于是有

$$\|Ax-b\|_2^2=\|Ax-b_1\|_2^2+\|b_2\|_2^2\geqslant\|b_2\|_2^2$$

由此，当

$$Ax=b_1=P_{R(A)}b$$

时，有

$$\|Ax-b\|_2=\min$$

由§7.1 定理3之 g)，当(15)式成立时，有

$$AA^{-}b=P_{R(A)}b=b_1$$

反之，若 $X\in C^{n\times m}$，则对任何 $b\in C^m$ 均使

$$\|AXb-b\|_2=\min$$

特别，依次选 $b=e_i \quad i\in \underline{m} \quad e_i=(0,\cdots,0,1,0,\cdots,0)^T$，则

$$AXe_i=P_{R(A)}e_i=AA^{-}e_i$$

且 $AA^{-}A=A$，$(AA^{-})^H=AA^{-}$，于是有

$$AX=AA^{-}$$

$$AXA=AA^{-}A=A$$

$$(AX)^H=(AA^{-})^H=AA^{-}=AX$$

定理8 设 $A\in C^{m\times n}$，$b\in C^m$，则

$$\|Ax-b\|_2=\min \tag{12}$$

的 LNLS 解是 A^+b；反之，若 $X\in C^{n\times m}$对任何 $b\in C^m$，Xb 都是(12)对应的 LNLS 解，则

$$X=A^+$$

证 由定理7，方程(12)的 LNLS 解，就是

$$Ax=AA^{-}b \quad (AA^{-}A=A,\ (AA^{-})^H=AA^{-})$$

的 LN 解，于是由定理6有

$$x=DAA^{-}b、ADA=A、(DA)^H=DA$$

易证

$$DAA^{-}=A^+$$

即

$$x=A^+b$$

反之，对任意的 $b\in C^m$，Xb 是(12)对应的 LNLS 解，A^+b 也是其 LNLS 解，则由解的唯一性有

$$Xb=A^+b$$

于是由 b 的任意性得

$$X=A^{+}$$

证毕。

作为本章的结束，对广义逆矩阵来说，在本节中已经用到未明确提出的满足矩阵方程

$$AXA=A、(AX)^{H}=AX$$

和满足矩阵方程

$$AXA=A、(XA)^{H}=XA$$

的广义逆矩阵。除此之外，还有其它很多有用的广义逆矩阵均因篇幅所限未能提及，而只对常用的几类作了简要介绍。

习　题　七

一、试证 $A^{H}(AA^{H})^{-}A$ 是 Hermite 阵，而且不论定成什么样的广义逆矩阵，$A^{H}(AA^{H})^{-}A$ 都不变。

二、令 $U=V(I_m-AA^{-})+(I_n-A^{-}A)W$，试证

$$X=A^{-}+V(I_m-AA^{-})+(I_n-A^{-}A)W$$

可写成

$$X=A^{-}+U-A^{-}AUAA^{-}$$

三、证 $O_{m\times n}$的自反广义逆矩阵仅为 $O_{n\times m}$。

四、设 $A\in C_r^{m\times n}$，$Y\in C^{n\times r}$，$Z\in C^{r\times m}$，且

$$ZAY=I_r$$

则 $A_r^{-}=YZ$ 是 A 的自反广义逆矩阵。

五、证明下列性质：

a）　当 $A=A^{H}$ 时，有

$$(A^2)^{+}=(A^{+})^2$$

$$AA^{+}=A^{-}A$$

$$A^2(A^2)^{+}=(A^2)^{+}A^2=AA^{+}$$

$$A^{+}A^2=A^2A^{+}$$

$b)$ $(A^HA)^+=A^+(A^H)^+$

$(AA^H)^+=(A^H)^+A^+$

$c)$ $(A^HA)^+=A^+(AA^H)^+A=A^H(AA^H)^+(A^H)^+$

$d)$ $AA^+=(AA^H)(AA^H)^+=(AA^H)^+(AA^H)$

六、若 $ABA=A$、$(BA)^H=BA$、$ACA=A$、$(AC)^H=AC$，则

$$BAC=A^+$$

七、证明$(A\otimes B)^+=A^+\otimes B^+$

八、证明

$$AA^+=Q^HQ$$

$$A^+A=PP^H$$

其中 $A=PDQ^H$，$D=\mathrm{diag}(\Sigma,O)$，$\Sigma=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$，$\sigma_i(i\in\underline{r})$为 A 的正奇值，$r=\mathrm{rank}A$。

九、设 $AA^H=A^HA=I$，$\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，则

$$(A\Lambda A)^+=A\Lambda^+A^H$$

但 A 只是可逆时，一般此式不成立，试举反例。

十、设

$$A=\begin{bmatrix}1 & 2 & -1\\ 0 & -1 & 2\end{bmatrix}$$

求 A^+。

十一、求

$$A=\begin{bmatrix}1 & 2 & 0\\ 0 & 0 & 2\\ 2 & 4 & 0\end{bmatrix}$$

的广义逆矩阵 A^-。

参 考 书 目

〔1〕黄 琳编著，《系统与控制理论中的线性代数》，科学出版社，1984。

〔2〕北京大学数学力学系几何与代数教研室代数小组编，《高等代数》，人民教育出版社，1978。

〔3〕（日）须田信英等著，曹长修译，《自动控制中的矩阵理论》，科学出版社，1979。

〔4〕Φ·P·甘特马赫尔著，柯召译，《矩阵论》（上、下册），高等教育出版社，1955。

〔5〕谢绪恺，《现代控制理论基础》，辽宁人民出版社，1981。



重印说明

本书1988年4月正式出版后，一些兄弟院校将它选作教材或指定为报考工科博士生的参考书。鉴于此，决定重印。

考虑到工科数学课程指导委员会已对工科硕士生的《矩阵理论》课的教学基本要求的教学需要作了明确规定，本科《线性代数》课的教学改革尚无定论等原因，故本版不宜作内容的大变动，只作了较少变动，习题也作了少量调整，纠正了第一版中的一些印刷错误。

限于我们的水平，错误难免出现，敬请批评指正。

编　者

1994年9月于哈尔滨工业大学